GenCore version 5.1.7

OM protein - protein search, using sw model

Run on: February 28, 2006, 08:37:25 ; Search time 218.872 Seconds
(without alignments)
1282.772 Million cell updates/sec

Title: US-10-717-665A-44
Perfect score: 3349
Sequence: 1 MPQPSVSGMDPPPGDAFRSH.....TAEGNPTGGLVGLRIPTSKV 639

Scoring table: BLOSUM62
Gapop 10.0 , Gapext 0.5

Searched: 2443163 seqs, 439378781 residues

Total number of hits satisfying chosen parameters: 2443163

Minimum DB seq length: 0
Maximum DB seq length: 2000000000

Post-processing: Minimum Match 0%
Maximum Match 100%
Listing first 45 summaries

Database : A_Geneseq_21:*
1: geneseqp1980s:*
2: geneseqp1990s:*
3: geneseqp2000s:*
4: geneseqp2001s:*
5: geneseqp2002s:*
6: geneseqp2003as:*
7: geneseqp2003bs:*
8: geneseqp2004s:*
9: geneseqp2005s:*

Pred. No. is the number of results predicted by chance to have a score greater than or equal to the score of the result being printed, and is derived by analysis of the total score distribution.

SUMMARIES

Result No.	Score	% Query Match	Length	DB	ID	Description
1	3349	100.0	639	7	ADK65805	Angiogene
2	2148	64.1	417	8	ADQ66472	Novel hum
3	1534	45.8	296	8	ADO20373	Human PRO
4	1268	37.9	256	5	ADK35935	Novel hum
5	865	25.8	168	5	ABP64928	Human pro
6	506	15.1	814	4	ABB68374	Drosophil
7	434	13.0	91	7	ADK65836	Angiogene
8	202	6.0	712	7	ADE61623	Rat Prote
9	202	6.0	712	7	ADD46090	Rat Prote
10	202	6.0	712	7	ADE57828	Rat Prote
11	202	6.0	712	7	ADE57830	Rat Prote
12	200.5	6.0	160	4	ABG29275	Novel hum
13	196.5	5.9	1162	3	AAY96255	Kaposi's
14	196.5	5.9	1162	3	AAY58500	HHV8 ORF
15	196.5	5.9	1162	4	AAB62331	Amino aci
16	196.5	5.9	1162	5	ABB05621	Kaposi's
17	196.5	5.9	1162	8	ADJ65096	HHV8 late
18	196.5	5.9	1162	9	ADV68154	Kaposi's
19	189.5	5.7	735	6	AAE37016	Human nuc
20	184	5.5	764	2	AAW01897	Nonsense-
21	184	5.5	764	3	AAY98056	Yeast NMD
22	184	5.5	1089	2	AAW01896	Nonsense-
23	184	5.5	1089	2	AAY05835	Yeast Nmd
24	184	5.5	1089	3	AAY98055	Yeast Nmd
25	181	5.4	842	4	AAB85725	Aab85725 Rat TBP-b
26	179	5.3	427	5	ABP73549	Abp73549 Candida a
27	179	5.3	3553	9	AEB22174	Aeb22174 Plasmodiu
28	178.5	5.3	1279	4	AAM39101	Aam39101 Human pol
29	178.5	5.3	1305	4	AAM40887	Aam40887 Human pol
30	175.5	5.2	709	7	ADD01163	Add01163 Human nuc
31	175	5.2	2724	6	ABP96961	Abp96961 Human BMC
32	174	5.2	520	8	ADP46649	Adp46649 Human col
33	174	5.2	568	8	ADP46648	Adp46648 Human col
34	174	5.2	706	7	ADE61625	Ade61625 Human Pro
35	174	5.2	706	7	ADD46092	Add46092 Human Pro
36	174	5.2	707	2	AAR79912	Aar79912 Human nuc
37	174	5.2	707	2	AAW84052	Aaw84052 Human V3
38	174	5.2	707	4	AAB48964	Aab48964 Human nuc
39	174	5.2	707	7	ADD49220	Add49220 Human nuc
40	174	5.2	707	8	ADJ58974	Adj58974 Human nuc
41	174	5.2	707	8	ADP54086	Adp54086 Human PRO
42	174	5.2	707	9	ADY19864	Ady19864 PRO polyp
43	174	5.2	710	8	ABM80398	Abm80398 Tumour-as
44	173.5	5.2	2097	4	ABG09944	Abg09944 Novel hum
45	172.5	5.2	722	7	ABM87025	Abm87025 Rice abio

ALIGNMENTS

RESULT 1
ADK65805
ID ADK65805 standard; protein; 639 AA.
XX
AC ADK65805;
XX
DT 06-MAY-2004 (first entry)
XX
DE Angiogenesis-differentially expressed protein ANH0757.
XX
KW cytostatic; cardiant; vasotropic; antiarteriosclerotic;
KW angiogenesis inhibitor; angiogenesis stimulator; angiogenic index;
KW gene expression; cancer; coronary artery disease; myocardial ischemia;
KW coronary arteriosclerosis; forensic medicine.
XX
OS Homo sapiens.
XX
PN WO2003066831-A2.
XX
PD 14-AUG-2003.
XX
PF 07-FEB-2003; 2003WO-US003848.
XX
PR 07-FEB-2002; 2002US-00067482.
PR 10-JUN-2002; 2002US-00164595.
PR 16-AUG-2002; 2002US-0403649P.
PR 03-JAN-2003; 2003US-0437746P.
XX
PA (ORIG-) ORIGENE TECHNOLOGIES INC.
XX
PI Sun Z, Li X, Kovacs KF, Fan W, Jay G;
XX
DR WPI; 2003-731502/69.
DR N-PSDB; ADK65804.
XX
PT Determining the angiogenic index of a tissue or cell sample using
PT expression levels of differentially expressed genes, useful for
PT diagnosing or treating cancer, coronary artery disease, myocardial
PT ischemia and/or arteriosclerosis.
XX
PS Claim 23; SEQ ID NO 44; 296pp; English.
XX
CC The invention relates to a method of determining the angiogenic index of
CC a tissue or cell sample comprising assessing, in a sample, the expression
CC levels of one or more differentially-expressed gene from any of 34 DNA
CC sequences, given in the specification, where the levels are indicative of
CC the angiogenic index. The methods and compositions of the present

```
CC   invention are useful for diagnosing, preventing and/or treating cancer,
CC   coronary artery disease, myocardial ischemia or coronary
CC   arteriosclerosis. They can also be used in research, drug discovery and
CC   forensic medicine involving angiogenesis. This sequence corresponds to
CC   one of the differentially expressed proteins of the invention.
XX
SQ   Sequence 639 AA;

  Query Match             100.0%;  Score 3349;  DB 7;  Length 639;
  Best Local Similarity   100.0%;  Pred. No. 8.4e-248;
  Matches  639;  Conservative    0;  Mismatches    0;  Indels    0;  Gaps    0;

Qy          1 MPQPSVSGMDPPFGDAFRSHTFSEQTLMSTDLLANSSDPDFMYELDREMNYQQNPRDNFL 60
              ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db          1 MPQPSVSGMDPPFGDAFRSHTFSEQTLMSTDLLANSSDPDFMYELDREMNYQQNPRDNFL 60

Qy         61 SLEDCKDIENLESFTDVLDNEGALTSNWEQWDTYCEDLTKYTKLTSCDIWGTKEVDYLGL 120
              ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db         61 SLEDCKDIENLESFTDVLDNEGALTSNWEQWDTYCEDLTKYTKLTSCDIWGTKEVDYLGL 120

Qy        121 DDFSSPYQDEEVISKTPTLAQLNSEDSQSVSDSLYYPDSLFSVKQNPLPSSFPGKKITSR 180
              ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db        121 DDFSSPYQDEEVISKTPTLAQLNSEDSQSVSDSLYYPDSLFSVKQNPLPSSFPGKKITSR 180

Qy        181 AAAPVCSSKTLQAEVPLSDCVQKASKPPSSTQIMVKTNMYHNEKVNFHVECKDYVKKAKV 240
              ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db        181 AAAPVCSSKTLQAEVPLSDCVQKASKPPSSTQIMVKTNMYHNEKVNFHVECKDYVKKAKV 240

Qy        241 KINPVQQSRPLLSQIHTDAAKENTCYCGAVAKRQEKKGMEPLQGHATPALPFKETQELLL 300
              ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db        241 KINPVQQSRPLLSQIHTDAAKENTCYCGAVAKRQEKKGMEPLQGHATPALPFKETQELLL 300

Qy        301 SPLPQEGPGSLAAGESSSLSASTSVSDSSQKKEEHNYSLFVSDNLGEQPTKCSPEEDEED 360
              ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db        301 SPLPQEGPGSLAAGESSSLSASTSVSDSSQKKEEHNYSLFVSDNLGEQPTKCSPEEDEED 360

Qy        361 EEDVDDEDHDEGFGSEHELSENEEEEEEEEDYEDDKDDDISDTFSEPGYENDSVEDLKEV 420
              ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db        361 EEDVDDEDHDEGFGSEHELSENEEEEEEEEDYEDDKDDDISDTFSEPGYENDSVEDLKEV 420

Qy        421 TSISSRKRGKRRYFWEYSEQLTPSQQERMLRPSEWNRDTLPSNMYQKNGLHHGKYAVKKS 480
              ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db        421 TSISSRKRGKRRYFWEYSEQLTPSQQERMLRPSEWNRDTLPSNMYQKNGLHHGKYAVKKS 480

Qy        481 RRTDVEDLTPNPKKLLQIGNELRKLNKVISDLTPVSELPLTARPRSRKEKNKLAFRACRL 540
              ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db        481 RRTDVEDLTPNPKKLLQIGNELRKLNKVISDLTPVSELPLTARPRSRKEKNKLAFRACRL 540

Qy        541 KKKAQYEANKVKLWGLNTEYDNLLFVINSIKQEIVNRVQNPRDERGPNMGQKLEILIKDT 600
              ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db        541 KKKAQYEANKVKLWGLNTEYDNLLFVINSIKQEIVNRVQNPRDERGPNMGQKLEILIKDT 600

Qy        601 LGLPVAGQTSEFVNQVLEKTAEGNPTGGLVGLRIPTSKV 639
              |||||||||||||||||||||||||||||||||||||||
Db        601 LGLPVAGQTSEFVNQVLEKTAEGNPTGGLVGLRIPTSKV 639


RESULT 2
ADQ66472
ID   ADQ66472 standard; protein; 417 AA.
XX
AC   ADQ66472;
XX
DT   07-OCT-2004  (first entry)
XX
DE   Novel human protein sequence #1445.
XX
KW   osteopathic; neuroprotective; nootropic; antiparkinsonian; cytostatic;
KW   gene therapy; diagnostic marker; morbid state; osteoporosis;
KW   neurological disease; Alzheimer's disease; Parkinson's disease; dementia;
KW   cancer.
XX
OS   Homo sapiens.
XX
PN   EP1440981-A2.
XX
PD   28-JUL-2004.
XX
PF   21-JAN-2004; 2004EP-00001196.
XX
PR   21-JAN-2003; 2003JP-00102206.
PR   09-MAY-2003; 2003JP-00131392.
XX
PA   (REAS-) RES ASSOC BIOTECHNOLOGY.
XX
PI   Isogai T,  Sugiyama T,  Otsuki T,  Wakamatsu A,  Sato H,  Ishii S;
PI   Yamamoto J,  Isono Y,  Nagai K,  Irie R;
XX
DR   WPI; 2004-535376/52.
DR   N-PSDB; ADQ64284.
XX
PT   Novel 2495 cDNA, useful for treating osteoporosis, neurological diseases,
PT   Alzheimer's diseases, Parkinson's diseases, dementia and various cancers.
XX
PS   Claim 1; SEQ ID NO 3633; 2449pp; English.
XX
CC   The invention relates to 2495 novel polynucleotides (I) and their encoded
CC   polypeptides, sequences hybridizing to these nucleotides, sequences
CC   encoding partial polypeptides and sequences having 70% or 90% identity to
CC   the nucleotide and protein sequences. The nucleotides and polypeptides
CC   are useful as diagnostic markers or therapeutic target for the diseases
CC   or morbid states. They are also useful for treating osteoporosis,
CC   neurological diseases, Alzheimer's diseases, Parkinson's diseases,
CC   dementia and various cancers. This sequence corresponds to a protein
CC   sequence of the invention.
XX
SQ   Sequence 417 AA;

  Query Match             64.1%;  Score 2148;  DB 8;  Length 417;
  Best Local Similarity   99.8%;  Pred. No. 6.5e-156;
  Matches  407;  Conservative    0;  Mismatches    1;  Indels    0;  Gaps    0;

Qy          1 MPQPSVSGMDPPFGDAFRSHTFSEQTLMSTDLLANSSDPDFMYELDREMNYQQNPRDNFL 60
              ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db          1 MPQPSVSGMDPPFGDAFRSHTFSEQTLMSTDLLANSSDPDFMYELDREMNYQQNPRDNFL 60

Qy         61 SLEDCKDIENLESFTDVLDNEGALTSNWEQWDTYCEDLTKYTKLTSCDIWGTKEVDYLGL 120
              ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db         61 SLEDCKDIENLESFTDVLDNEGALTSNWEQWDTYCEDLTKYTKLTSCDIWGTKEVDYLGL 120

Qy        121 DDFSSPYQDEEVISKTPTLAQLNSEDSQSVSDSLYYPDSLFSVKQNPLPSSFPGKKITSR 180
              ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db        121 DDFSSPYQDEEVISKTPTLAQLNSEDSQSVSDSLYYPDSLFSVKQNPLPSSFPGKKITSR 180

Qy        181 AAAPVCSSKTLQAEVPLSDCVQKASKPPSSTQIMVKTNMYHNEKVNFHVECKDYVKKAKV 240
              |||||||||||||||||||||||||| ||||||||||||||||||||||||||||||||||
Db        181 AAAPVCSSKTLQAEVPLSDCVQKASKPTSSTQIMVKTNMYHNEKVNFHVECKDYVKKAKV 240

Qy        241 KINPVQQSRPLLSQIHTDAAKENTCYCGAVAKRQEKKGMEPLQGHATPALPFKETQELLL 300
              ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db        241 KINPVQQSRPLLSQIHTDAAKENTCYCGAVAKRQEKKGMEPLQGHATPALPFKETQELLL 300

Qy        301 SPLPQEGPGSLAAGESSSLSASTSVSDSSQKKEEHNYSLFVSDNLGEQPTKCSPEEDEED 360
              ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db        301 SPLPQEGPGSLAAGESSSLSASTSVSDSSQKKEEHNYSLFVSDNLGEQPTKCSPEEDEED 360

Qy        361 EEDVDDEDHDEGFGSEHELSENEEEEEEEEDYEDDKDDDISDTFSEPG 408
              |||||||||||||||||||||||||||||||||||||||||||||||||
Db        361 EEDVDDEDHDEGFGSEHELSENEEEEEEEEDYEDDKDDDISDTFSEPG 408


RESULT 3
ADO20373
ID   ADO20373 standard; protein; 296 AA.
XX
AC   ADO20373;
```

```
XX
DT   12-AUG-2004  (first entry)
XX
DE   Human PRO polypeptide #633.
XX
KW   Human; PRO; immune related disorder; systemic lupus erythematosus;
KW   rheumatoid arthritis; osteoarthritis; juvenile chronic arthritis;
KW   systemic sclerosis; Sjogren's syndrome; vasculitis; sarcoidosis;
KW   autoimmune haemolytic anaemia; autoimmune thrombocytopenia; thyroiditis;
KW   diabetes mellitus; renal disease; demyelinating disease;
KW   central nervous system; peripheral nervous system;
KW   demyelinating polyneuropathy; Guillain-Barre syndrome;
KW   chronic inflammatory demyelinating polyneuropathy.
XX
OS   Homo sapiens.
XX
PN   WO2004043361-A2.
XX
PD   27-MAY-2004.
XX
PF   06-NOV-2003; 2003WO-US035268.
XX
PR   08-NOV-2002; 2002US-0425235P.
XX
PA   (GETH ) GENENTECH INC.
XX
PI   Fong S,  Dennis K,  Clark H,  Chiu H,  Schoenfeld J,  Williams PM;
PI   Wood WI,  Wu TD;
XX
DR   WPI; 2004-420067/39.
DR   N-PSDB; ADO20372.
XX
PT   Novel PRO polypeptide e.g., PRO69614, PRO71106, or PRO86388 useful for
PT   treating an immune related disorder such as systemic lupus erythematosus,
PT   rheumatoid arthritis, osteoarthritis, juvenile chronic arthritis or
PT   spondyloarthropathy.
XX
PS   Claim 7; SEQ ID NO 1266; 1731pp; English.
XX
CC   The invention relates to human PRO polypeptides and the polynucleotides
CC   encoding them. The polypeptides and polynucleotides are useful for
CC   treating and diagnosing immune related disorders in mammals. The immune
CC   related disorders include systemic lupus erythematosus, rheumatoid
CC   arthritis, osteoarthritis, juvenile chronic arthritis, systemic
CC   sclerosis, Sjogren's syndrome, vasculitis, sarcoidosis, autoimmune
CC   haemolytic anaemia, autoimmune thrombocytopenia, thyroiditis, diabetes
CC   mellitus, immune-mediated renal disease, demyelinating diseases of the
CC   central or peripheral nervous system, demyelinating polyneuropathy,
CC   Guillain-Barre syndrome and chronic inflammatory demyelinating
CC   polyneuropathy. This sequence represents a human PRO polypeptide of the
CC   invention.
XX
SQ   Sequence 296 AA;

  Query Match             45.8%;  Score 1534;  DB 8;  Length 296;
  Best Local Similarity   99.3%;  Pred. No. 5.1e-109;
  Matches  294;  Conservative    0;  Mismatches    2;  Indels    0;  Gaps    0;

Qy        344 NLGEQPTKCSPEEDEEDEEDVDDEDHDEGFGSEHELSENEEEEEEEEDYEDDKDDDISDT 403
              ||||||||||||||||||||||||||||||||||||||||||||||| ||||||||||||||
Db          1 NLGEQPTKCSPEEDEEDEEDVDDEDHDEGFGSEHELSENEEEEEVEEDYEDDKDDDISDT 60

Qy        404 FSEPGYENDSVEDLKEVTSISSRKRGKRRYFWEYSEQLTPSQQERMLRPSEWNRDTLPSN 463
              ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db         61 FSEPGYENDSVEDLKEVTSISSRKRGKRRYFWEYSEQLTPSQQERMLRPSEWNRDTLPSN 120

Qy        464 MYQKNGLHHGKYAVKKSRRTDVEDLTPNPKKLLQIGNELRKLNKVISDLTPVSELPLTAR 523
              ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db        121 MYQKNGLHHGKYAVKKSRRTDVEDLTPNPKKLLQIGNELRKLNKVISDLTPVSELPLTAR 180

Qy        524 PRSRKEKNKLAFRACRLKKKAQYEANKVKLWGLNTEYDNLLFVINSIKQEIVNRVQNPRD 583
              |||||||||| |||||||||||||||||||||||||||||||||||||||||||||||||||
Db        181 PRSRKEKNKLASRACRLKKKAQYEANKVKLWGLNTEYDNLLFVINSIKQEIVNRVQNPRD 240

Qy        584 ERGPNMGQKLEILIKDTLGLPVAGQTSEFVNQVLEKTAEGNPTGGLVGLRIPTSKV 639
              ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db        241 ERGPNMGQKLEILIKDTLGLPVAGQTSEFVNQVLEKTAEGNPTGGLVGLRIPTSKV 296


RESULT 4
ADK35935
ID   ADK35935 standard; protein; 256 AA.
XX
AC   ADK35935;
XX
DT   06-MAY-2004  (first entry)
XX
DE   Novel human polypeptide SeqID8017.
XX
KW   antiarthritic; antiparkinsonian; neuroprotective; nootropic;
KW   immunosuppressive; cytostatic; antipsoriatic; antiinflammatory;
KW   antibacterial; antiviral; antifungal; antiparasitic; gene therapy;
KW   arthritis; Parkinson's; Alzheimer's; autoimmune disease; cancer;
KW   psoriasis; inflammatory bowel disease; infection; bacteria; virus;
KW   fungus; parasite; human.
XX
OS   Homo sapiens.
XX
FH   Key             Location/Qualifiers
FT   Misc-difference 1. .256
FT                   /label= OTHER
FT                   /note= "OTHER= All Xaa's in this sequence are unknown
FT                   amino acids or the site of a stop codon within the DNA
FT                   sequence"
XX
PN   WO200216439-A2.
XX
PD   28-FEB-2002.
XX
PF   05-MAR-2001; 2001WO-US004941.
XX
PR   07-MAR-2000; 2000US-00519705.
PR   19-MAY-2000; 2000US-00574454.
XX
PA   (HYSE-) HYSEQ INC.
XX
PI   Tang YT,  Liu C,  Drmanac RT;
XX
DR   WPI; 2002-280918/32.
XX
PT   Isolated polynucleotide encoding bone marrow derived polypeptides useful
PT   for treating, e.g., Parkinson's, Alzheimer's, cancer, arthritis, Crohn's
PT   disease, and inflammatory bowel disease.
XX
PS   Claim 20; SEQ ID NO 8017; 504pp; English.
XX
CC   This invention relates to a novel isolated polynucleotide comprising a
CC   nucleotide sequence selected from one of 1680 sequences, a mature protein
CC   coding portion of them, an active domain of them and their complementary
CC   sequences. The invention may be useful for the production of compounds
CC   with an antiarthritic, antiparkinsonian, neuroprotective, nootropic,
CC   immunosuppressive, cytostatic, antipsoriatic, antiinflammatory,
CC   antibacterial, antiviral, antifungal or antiparasitic activity. In
CC   addition, the disclosed sequences may be useful for gene therapy. The
CC   polypeptides or their antibodies are useful for treating many diseases
CC   such as arthritis, Parkinson's, Alzheimer's, autoimmune diseases, cancer,
CC   psoriasis, inflammatory bowel disease and infections caused by bacteria,
CC   viruses, fungi or parasites. The present sequence is that of a human
CC   polypeptide of the invention.
XX
SQ   Sequence 256 AA;

  Query Match             37.9%;  Score 1268;  DB 5;  Length 256;
  Best Local Similarity   98.8%;  Pred. No. 1.1e-88;
  Matches  243;  Conservative    0;  Mismatches    3;  Indels    0;  Gaps    0;
```

```
Qy        163 VKQNPLPSSFPGKKITSRAAAPVCSSKTLQAEVPLSDCVQKASKPPSSTQIMVKTNMYHN 222
              ||||||||||||||| | ||||||||||||||||||||||||| ||||||||||||||
Db          5 VKQNPLPSSFPGKKITIRLAAPVCSSKTLQAEVPLSDCVQKASKPTSSTQIMVKTNMYHN 64

Qy        223 EKVNFHVECKDYVKKAKVKINPVQQSRPLLSQIHTDAAKENTCYCGAVAKRQEKKGMEPL 282
              ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db         65 EKVNFHVECKDYVKKAKVKINPVQQSRPLLSQIHTDAAKENTCYCGAVAKRQEKKGMEPL 124

Qy        283 QGHATPALPFKETQELLLSPLPQEGPGSLAAGESSSLSASTSVSDSSQKKEEHNYSLFVS 342
              ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db        125 QGHATPALPFKETQELLLSPLPQEGPGSLAAGESSSLSASTSVSDSSQKKEEHNYSLFVS 184

Qy        343 DNLGEQPTKCSPEEDEEDEEDVDDEDHDEGFGSEHELSENEEEEEEEEDYEDDKDDDISD 402
              ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db        185 DNLGEQPTKCSPEEDEEDEEDVDDEDHDEGFGSEHELSENEEEEEEEEDYEDDKDDDISD 244

Qy        403 TFSEPG 408
              ||||||
Db        245 TFSEPG 250
```

RESULT 5
ABP64928
ID ABP64928 standard; protein; 168 AA.
XX
AC ABP64928;
XX
DT 25-FEB-2003 (first entry)
XX
DE Human protein SEQ ID 588.
XX
KW Human; expressed sequence tag; EST; haematopoietic disorder;
KW central nervous system disease; viral infection;
KW peripheral nervous system disease; non-healing wound; infectious disease;
KW immune deficiency; immune disorder; bacterial infection; allergy; cancer;
KW fungal infection; autoimmune disorder; coagulation disorder; nootropic;
KW antiallergic; antiinflammatory; immunosuppressive; neuroprotective;
KW cytostatic; haemostatic; virucide; antibacterial; fungicide;
KW immunostimulant; cerebroprotective.
XX
OS Homo sapiens.
XX
PN WO200259260-A2.
XX
PD 01-AUG-2002.
XX
PF 16-NOV-2001; 2001WO-US042950.
XX
PR 17-NOV-2000; 2000US-00714936.
XX
PA (HYSE-) HYSEQ INC.
XX
PI Tang YT, Goodrich RW, Liu C, Zhou P, Asundi V, Zhang J, Zhao QA;
PI Ren F, Xue AJ, Yang Y, Wehrman T, Drmanac RT;
XX
DR WPI; 2002-590824/63.
DR N-PSDB; ABQ99514.
XX
PT New isolated polynucleotide, useful in research, diagnostic or
PT therapeutic methods, e.g. preventing or treating disorders involving
PT aberrant protein expression or biological activity.
XX
PS Claim 20; SEQ ID NO 588; 394pp; English.
XX
CC The present invention relates to novel human coding sequences (ABQ99268-
CC ABQ99608) and proteins (ABP64682-ABP65022). The sequences are useful in
CC therapeutic, diagnostic and research methods. The polynucleotides may be
CC used in the field of molecular biology as hybridisation probes, primers
CC for PCR, for chromosome and gene mapping, for the recombinant production
CC of protein, or in generation of anti-sense DNA or RNA. The
CC polynucleotides are useful in diagnostics as expressed sequence tags
CC (ESTs) for identifying expressed genes or for physical mapping of the
CC human genome. The proteins may be used as molecular weight markers, or as
CC nutritional sources or supplements. The proteins may be used to maintain
CC and expand cell population in a totipotential or pluripotential state
CC useful for re-engineering damaged or diseased tissues, transplantation,
CC manufacture of bio-pharmaceuticals or the development of bio-sensors. The
CC polynucleotides and proteins are useful for preventing, treating or
CC ameliorating disorders involving aberrant protein expression or
CC biological activity, e.g. haematopoietic disorders, central/peripheral
CC nervous system diseases, mechanical and traumatic disorders, non-healing
CC wounds, immune deficiencies and disorders, infectious diseases caused by
CC viral, bacterial or fungal infection, autoimmune disorders, allergic
CC reactions and conditions, coagulation disorders, or cancer. The
CC polynucleotide sequences of the invention were assembled from ESTs
CC isolated mainly by sequencing by hybridisation, and in some cases,
CC sequences obtained from one or more public databases. Note: The sequence
CC data for this patent did not form part of the printed specification, but
CC was obtained in electronic format directly from WIPO at
CC ftp.wipo.int/pub/published_pct_sequences
XX
SQ Sequence 168 AA;

```
  Query Match             25.8%;  Score 865;  DB 5;  Length 168;
  Best Local Similarity   100.0%;  Pred. No. 4.9e-58;
  Matches  164;  Conservative    0;  Mismatches    0;  Indels    0;  Gaps    0;

Qy        214 MVKTNMYHNEKVNFHVECKDYVKKAKVKINPVQQSRPLLSQIHTDAAKENTCYCGAVAKR 273
              ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db          1 MVKTNMYHNEKVNFHVECKDYVKKAKVKINPVQQSRPLLSQIHTDAAKENTCYCGAVAKR 60

Qy        274 QEKKGMEPLQGHATPALPFKETQELLLSPLPQEGPGSLAAGESSSLSASTSVSDSSQKKE 333
              ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db         61 QEKKGMEPLQGHATPALPFKETQELLLSPLPQEGPGSLAAGESSSLSASTSVSDSSQKKE 120

Qy        334 EHNYSLFVSDNLGEQPTKCSPEEDEEDEEDVDDEDHDEGFGSEH 377
              |||||||||||||||||||||||||||||||||||||||||||||
Db        121 EHNYSLFVSDNLGEQPTKCSPEEDEEDEEDVDDEDHDEGFGSEH 164
```

RESULT 6
ABB68374
ID ABB68374 standard; protein; 814 AA.
XX
AC ABB68374;
XX
DT 26-MAR-2002 (first entry)
XX
DE Drosophila melanogaster polypeptide SEQ ID NO 31914.
XX
KW Drosophila; developmental biology; cell signalling; insecticide;
KW pharmaceutical.
XX
OS Drosophila melanogaster.
XX
PN WO200171042-A2.
XX
PD 27-SEP-2001.
XX
PF 23-MAR-2001; 2001WO-US009231.
XX
PR 23-MAR-2000; 2000US-0191637P.
PR 11-JUL-2000; 2000US-00614150.
XX
PA (PEKE) PE CORP NY.
XX
PI Venter JC, Adams M, Li PWD, Myers EW;
XX
DR WPI; 2001-656860/75.
DR N-PSDB; ABL12477.
XX
PT New isolated nucleic acid detection reagent for detecting 1000 or more
PT genes from Drosophila and for elucidating cell signaling and cell-cell
PT interactions.
XX
PS Disclosure; SEQ ID NO 31914; 21pp + Sequence Listing; English.

```
XX
CC   The invention relates to an isolated nucleic acid detection reagent
CC   capable of detecting 1000 or more genes from Drosophila. The invention is
CC   useful in developmental biology and in elucidating cell signalling and
CC   cell-cell interactions in higher eukaryotes for the development of
CC   insecticides, therapeutics and pharmaceutical drugs. The invention
CC   discloses genomic DNA sequences (ABL16176-ABL30511), expressed DNA
CC   sequences (ABL01840-ABL16175) and the encoded proteins (ABB57737-
CC   ABB72072). The sequence data for this patent did not form part of the
CC   printed specification, but was obtained in electronic format directly
CC   from WIPO at ftp.wipo.int/pub/published_pct_sequences
XX
SQ   Sequence 814 AA;

  Query Match             15.1%;  Score 506;  DB 4;  Length 814;
  Best Local Similarity   25.0%;  Pred. No. 1.7e-29;
  Matches  186;  Conservative   94;  Mismatches  211;  Indels  254;  Gaps   25;

Qy          5 SVSGMDPPFGDAF---RSHTFSEQTLMSTDLLANSSDPDFMYELDREMNYQQNPRDNFL- 60
Db        205 SSNGIGSGSGGYFTPDMSHSLSLNVVSEQVLLQEATTPN-------ELLYEMTPNSNAMW 257

Qy         61 ---------------SLED------------CKDIENLE--SFTDVLDNEGALTSNWEQW 91
Db        258 SDISSAIIHTKHEPFSLDDDYIFPNDKAEIQAADLSDLNGGDFLDVIGN---------- 306

Qy         92 DTYCEDLTKYTKLTSCDIWGTKEVDYL------GLDDFSSP----------YQDEEV-- 132
Db        307 ---IEDFLPQTAV-------TQSVNFLLSPQAQGQDALVAPPMELLQQQQQNHQQLQVGS 356

Qy        133 ---ISKTPTLAQLNSEDSQSVSDSLYYPDSLFSVKQNPLPSSF-PGKKITSRAAAPVCSS 188
Db        357 LPQLQTLLTLSQQQQSNSSSTSPYEIYHSTPQKPQQQQLSASFSPGSQASQSPLTP---- 412

Qy        189 KTLQAEVPLSDCVQKASKPPSST----QIMVKTNMYHNEKVNFHVECKDYVKKAKVKINP 244
Db        413 ---------------PPPPHANRPQYQMVKSRNM-----------QELIKKGFPMSSP 444

Qy        245 VQQSRPLLSQIHTDAAKENTCYCGAVAKRQEKKGMEPLQGHATPALPFKETQELLLSPLP 304
Db        445 PERS--ILS-----------------------------------------QSAALSPGG 460

Qy        305 QEGPGSLAAGESSSLSASTSVSDSSQKKEEHNYSLFVSDNLGEQPTKCSPEEDEEDEEDV 364
Db        461 SSGFGSSASGNSTTTSNQTS---GSAVRKSFGYQSAVENSQLSRLSSSAPTH-------- 509

Qy        365 DDEDHDEGFGSEH---------------ELSENEE----------------EEEEEEDY 392
Db        510 --------LGLEHIWMRREPRQHLLSTGSLAEAESFSSLSTGSVLSPDGIDFSQDDEDDN 561

Qy        393 EDDKDDDISDTFSEPGYENDSVEDLKEVTSISSRKRGKRRYFWEYSEQLTPSQQERML-- 450
Db        562 SSENSDNYDDCSSDNGLSEDEDETRTSTPNHLSSSKGKERFFWQYNVQAKGPKGKRLVFQ 621

Qy        451 ----RPSEWNRDTLP----------SNMYQKNGLHHGKYAVKKSRRTDVEDLTPNPKKLL 496
Db        622 SKLEDPHVLNEVTDPVFSPTCSVRGIKVYKHSG---------KARKGDGNDLTPNARKLH 672

Qy        497 QIGNELRKLNKVISDLTPVSELPLTARPRSRKEKNKLAFRACRLKKKAQYEANKVKLWGL 556
Db        673 NIGKELDKLSRTINDMTPVSELPFNVRPKSRKEKNKLASRACRLKKKAQHEANKIKLFGL 732

Qy        557 NTEYDNLLFVINSIKQEIVNRVQNPRDERGPNMGQKLE--------ILIKDTLGLPVAGQ 608
Db        733 EIEHKRLMNGIAELKQALVVK------HRTKNLGESTEEVDQQIARIYATASSGIRIAGG 786

Qy        609 TSEFVNQVLEKTAEGNPTGGLVGLR 633
Db        787 STDFVNKVLENMRGGMPNGGLEELR 811


RESULT 7
ADK65836
ID   ADK65836 standard; protein; 91 AA.
XX
AC   ADK65836;
XX
DT   06-MAY-2004  (first entry)
XX
DE   Angiogenesis-differentially expressed protein #50.
XX
KW   cytostatic; cardiant; vasotropic; antiarteriosclerotic;
KW   angiogenesis inhibitor; angiogenesis stimulator; angiogenic index;
KW   gene expression; cancer; coronary artery disease; myocardial ischemia;
KW   coronary arteriosclerosis; forensic medicine.
XX
OS   Homo sapiens.
XX
PN   WO2003066831-A2.
XX
PD   14-AUG-2003.
XX
PF   07-FEB-2003; 2003WO-US003848.
XX
PR   07-FEB-2002; 2002US-00067482.
PR   10-JUN-2002; 2002US-00164595.
PR   16-AUG-2002; 2002US-0403649P.
PR   03-JAN-2003; 2003US-0437746P.
XX
PA   (ORIG-) ORIGENE TECHNOLOGIES INC.
XX
PI   Sun Z,  Li X,  Kovacs KF,  Fan W,  Jay G;
XX
DR   WPI; 2003-731502/69.
XX
PT   Determining the angiogenic index of a tissue or cell sample using
PT   expression levels of differentially expressed genes, useful for
PT   diagnosing or treating cancer, coronary artery disease, myocardial
PT   ischemia and/or arteriosclerosis.
XX
PS   Disclosure; SEQ ID NO 75; 296pp; English.
XX
CC   The invention relates to a method of determining the angiogenic index of
CC   a tissue or cell sample comprising assessing, in a sample, the expression
CC   levels of one or more differentially-expressed gene from any of 34 DNA
CC   sequences, given in the specification, where the levels are indicative of
CC   the angiogenic index. The methods and compositions of the present
CC   invention are useful for diagnosing, preventing and/or treating cancer,
CC   coronary artery disease, myocardial ischemia or coronary
CC   arteriosclerosis. They can also be used in research, drug discovery and
CC   forensic medicine involving angiogenesis. This sequence corresponds to
CC   one of the differentially expressed proteins of the invention.
XX
SQ   Sequence 91 AA;

  Query Match             13.0%;  Score 434;  DB 7;  Length 91;
  Best Local Similarity   100.0%;  Pred. No. 2.3e-25;
  Matches   83;  Conservative    0;  Mismatches    0;  Indels    0;  Gaps    0;

Qy        449 MLRPSEWNRDTLPSNMYQKNGLHHGKYAVKKSRRTDVEDLTPNPKKLLQIGNELRKLNKV 508
Db          1 MLRPSEWNRDTLPSNMYQKNGLHHGKYAVKKSRRTDVEDLTPNPKKLLQIGNELRKLNKV 60

Qy        509 ISDLTPVSELPLTARPRSRKEKN 531
Db         61 ISDLTPVSELPLTARPRSRKEKN 83


RESULT 8
ADE61623
ID   ADE61623 standard; protein; 712 AA.
XX
AC   ADE61623;
XX
DT   29-JAN-2004  (first entry)
XX
```

```
DE   Rat Protein P13383, SEQ ID NO 7544.
XX
KW   Rat; pain; neuronal tissue; gene therapy; spinal segmental nerve injury;
KW   chronic constriction injury; CCI; spared nerve injury; SNI; Chung.
XX
OS   Rattus norvegicus.
XX
PN   WO2003016475-A2.
XX
PD   27-FEB-2003.
XX
PF   14-AUG-2002; 2002WO-US025765.
XX
PR   14-AUG-2001; 2001US-0312147P.
PR   01-NOV-2001; 2001US-0346382P.
PR   26-NOV-2001; 2001US-0333347P.
XX
PA   (GEHO ) GEN HOSPITAL CORP.
PA   (FARB ) BAYER AG.
XX
PI   Woolf C,  D'urso D,  Befort K,  Costigan M;
XX
DR   WPI; 2003-268312/26.
DR   GENBANK; P13383.
XX
PT   New composition comprising two or more isolated polypeptides, useful for
PT   preparing a medicament for treating pain in an animal.
XX
PS   Claim 1; Page; 1017pp; English.
XX
CC   The invention discloses a composition comprising two or more isolated rat
CC   or human polynucleotides or a polynucleotide which represents a fragment,
CC   derivative or allelic variation of the nucleic acid sequence. Also
CC   claimed are a vector comprising the novel polynucleotide, a host cell
CC   comprising the vector, a method for identifying a nucleotide sequence
CC   which is differentially regulated in an animal subjected to pain and a
CC   kit to perform the method, an array, a method for identifying an agent
CC   that increases or decreases the expression of the polynucleotide sequence
CC   that is differentially expressed in neuronal tissue of a first animal
CC   subjected to pain, a method for identifying a compound which regulates
CC   the expression of a polynucleotide sequence which is differentially
CC   expressed in an animal subjected to pain, a method for identifying a
CC   compound that regulates the activity of one or more of the
CC   polynucleotides, a method for producing a pharmaceutical composition, a
CC   method for identifying a compound or small molecule that regulates the
CC   activity in an animal of one or more of the polypeptides given in the
CC   specification, a method for identifying a compound useful in treating
CC   pain and a pharmaceutical composition comprising the one or more
CC   polypeptides or their antibodies. The polynucleotide or the compound that
CC   modulates its activity is useful for preparing a medicament for treating
CC   pain (e.g. spinal segmental nerve injury (Chung), chronic constriction
CC   injury (CCI) and spared nerve injury (SNI)) in an animal (e.g. gene
CC   therapy). The sequence presented is a rat protein (shown in Table 2 of
CC   the specification) which is differentially expressed during pain. Note:
CC   The sequence data for this patent did not form part of the printed
CC   specification, but was obtained in electronic form directly from WIPO at
CC   ftp.wipo.int/pub/published_pct_sequences.
XX
SQ   Sequence 712 AA;

  Query Match             6.0%;  Score 202;  DB 7;  Length 712;
  Best Local Similarity   22.4%;  Pred. No. 3.1e-06;
  Matches  107;  Conservative   60;  Mismatches  171;  Indels  140;  Gaps   18;

Qy        129 DEEVISKTPTLAQLNSEDSQSVSDSLYYPDSLFSVKQNPLPSSFPGKKITSRAAAPVCSS 188
              : : ::  |   : :||| :   |     |      ::  :|    ||| |:  |  | |
Db         12 ESKKMAPPPKEVEEDSEDEEMSEDE----DDSSGEEEVVIPQK-KGKKATTTPAKKVVVS 66

Qy        189 KTLQAEVPLSDCVQKASKPPSSTQIMVKTNMYHNEKVNFHVECKDYVKKAKVKINPVQQS 248
              :| :| ||     :||:  |               |       |  |  |||   |
Db         67 QTKKAAVPTP--AKKAAVTPG-------------KKAAATPAKKAVTPAKVVPTP---- 106

Qy        249 RPLLSQIHTDAAKENTCYCGAVAKRQEKKGMEPLQGHATPALPFKETQELLLSPLPQEGP 308
```

```
                  |:      |:     |||        |||   |
Db        107 ----------GKKGAAQAKALVPTPGKKGA------VTPAKGAKN-------------- 135

Qy        309 GSLAAGESSSLSASTSVSDSSQKKEEHNYSLFVSDNLGEQPTKCSP------EEDEEDEE 362
              |  | | |          | | : |:           | :| | :|      :||||::
Db        136 GKNAKKEDSDEDEDEEDEDDSDEDEDEEDEFEPPVVKGVKPAKAAPAAPASEDEDEEDDD 195

Qy        363 DVDDEDHDEGFGSEHELSENEEE----------------------EEEEEDYEDDKDD 398
              | ||:| ||    | |  ::|||                      |||||  |||:|:
Db        196 DEDDDDDDE---EEEEEDDSEEEVMEITPAKGKKTPAKVVPVKAKSVAEEEEDDEDDEDE 252

Qy        399 DISDTFSEPGYENDSVEDLKEVTSISSRKRGKRRYFWEYSEQLTPSQQ--ERMLRPSEWN 456
              :  :   |   |:|  || :|         |||:      :::|  ::  |   :  | :
Db        253 EEDE--DEEDEEDDEDEDEEEEEEPVKAAPGKRK------KEMTKQKEAPEAKKQKIEGS 304

Qy        457 RDTLPSNMYQKNGLHHGKYAVKKSRRTDVEDLTPNPKKLLQIGNELRKLNKVISDLTPVS 516
              | | |::                  : :| ||          : :|   ||:|   :
Db        305 EPTTPFNLF-----------------IGNLNPN--------KSVAELKVAISELFAKN 337

Qy        517 EL-PLTARPRSRKEKNKLAFRACRLKKKAQYEANKVKLWGLNTEYDNLLFVINSIKQE 573
              :|  :  |  : ::   : | :   :||  |   :|::|              | || |
Db        338 DLAAVDVRTGTNRKFGYVDFESAEDLEKA-LELTGLKVFG-----------NEIKLE 382
```

```
RESULT 9
ADD46090
ID   ADD46090 standard; protein; 712 AA.
XX
AC   ADD46090;
XX
DT   29-JAN-2004  (first entry)
XX
DE   Rat Protein P13383, SEQ ID NO 11765.
XX
KW   Rat; pain; neuronal tissue; gene therapy; spinal segmental nerve injury;
KW   chronic constriction injury; CCI; spared nerve injury; SNI; Chung.
XX
OS   Rattus norvegicus.
XX
PN   WO2003016475-A2.
XX
PD   27-FEB-2003.
XX
PF   14-AUG-2002; 2002WO-US025765.
XX
PR   14-AUG-2001; 2001US-0312147P.
PR   01-NOV-2001; 2001US-0346382P.
PR   26-NOV-2001; 2001US-0333347P.
XX
PA   (GEHO ) GEN HOSPITAL CORP.
PA   (FARB ) BAYER AG.
XX
PI   Woolf C,  D'urso D,  Befort K,  Costigan M;
XX
DR   WPI; 2003-268312/26.
DR   GENBANK; P13383.
XX
PT   New composition comprising two or more isolated polypeptides, useful for
PT   preparing a medicament for treating pain in an animal.
XX
PS   Claim 1; Page; 1017pp; English.
XX
CC   The invention discloses a composition comprising two or more isolated rat
CC   or human polynucleotides or a polynucleotide which represents a fragment,
CC   derivative or allelic variation of the nucleic acid sequence. Also
CC   claimed are a vector comprising the novel polynucleotide, a host cell
CC   comprising the vector, a method for identifying a nucleotide sequence
CC   which is differentially regulated in an animal subjected to pain and a
CC   kit to perform the method, an array, a method for identifying an agent
CC   that increases or decreases the expression of the polynucleotide sequence
CC   that is differentially expressed in neuronal tissue of a first animal
CC   subjected to pain, a method for identifying a compound which regulates
CC   the expression of a polynucleotide sequence which is differentially
```

```
CC   expressed in an animal subjected to pain, a method for identifying a
CC   compound that regulates the activity of one or more of the
CC   polynucleotides, a method for producing a pharmaceutical composition, a
CC   method for identifying a compound or small molecule that regulates the
CC   activity in an animal of one or more of the polypeptides given in the
CC   specification, a method for identifying a compound useful in treating
CC   pain and a pharmaceutical composition comprising the one or more
CC   polypeptides or their antibodies. The polynucleotide or the compound that
CC   modulates its activity is useful for preparing a medicament for treating
CC   pain (e.g. spinal segmental nerve injury (Chung), chronic constriction
CC   injury (CCI) and spared nerve injury (SNI)) in an animal (e.g. gene
CC   therapy). The sequence presented is a rat protein (shown in Table 2 of
CC   the specification) which is differentially expressed during pain. Note:
CC   The sequence data for this patent did not form part of the printed
CC   specification, but was obtained in electronic form directly from WIPO at
CC   ftp.wipo.int/pub/published_pct_sequences.
XX
SQ   Sequence 712 AA;

  Query Match             6.0%;  Score 202;  DB 7;  Length 712;
  Best Local Similarity   22.4%;  Pred. No. 3.1e-06;
  Matches  107;  Conservative   60;  Mismatches  171;  Indels  140;  Gaps   18;

Qy        129 DEEVISKTPTLAQLNSEDSQSVSDSLYYPDSLFSVKQNPLPSSFPGKKITSRAAAPVCSS 188
              : : ::  |   : :||| :   |     |     ::  :|    ||| |:  |  |  |
Db         12 ESKKMAPPPKEVEEDSEDEEMSEDE----DDSSGEEEVVIPQK-KGKKATTTPAKKVVVS 66

Qy        189 KTLQAEVPLSDCVQKASKPPSSTQIMVKTNMYHNEKVNFHVECKDYVKKAKVKINPVQQS 248
              :| :| ||     :||:  |               |       |  |  |||   |
Db         67 QTKKAAVPTP--AKKAAVTPG-------------KKAAATPAKKAVTPAKVVPTP---- 106

Qy        249 RPLLSQIHTDAAKENTCYCGAVAKRQEKKGMEPLQGHATPALPFKETQELLLSPLPQEGP 308
                        |:     |:    |||       |||   |
Db        107 ----------GKKGAAQAKALVPTPGKKGA------VTPAKGAKN-------------- 135

Qy        309 GSLAAGESSSLSASTSVSDSSQKKEEHNYSLFVSDNLGEQPTKCSP------EEDEEDEE 362
              |  |  | |          | | : |:           | :| | :|      :|||||::
Db        136 GKNAKKEDSDEDEDEEDEDDSDEDEDEEDEFEPPVVKGVKPAKAAPAAPASEDEDEEDDD 195

Qy        363 DVDDEDHDEGFGSEHELSENEEE----------------------EEEEEDYEDDKDD 398
              | ||:| ||    | |  ::|||                        ||||| |||:|:
Db        196 DEDDDDDDE---EEEEEDDSEEEVMEITPAKGKKTPAKVVPVKAKSVAEEEEDDEDDEDE 252

Qy        399 DISDTFSEPGYENDSVEDLKEVTSISSRKRGKRRYFWEYSEQLTPSQQ--ERMLRPSEWN 456
              :  :   |   |:|  || :|         |||:      :::|  ::  |   :  | :
Db        253 EEDE--DEEDEEDDEDEDEEEEEEPVKAAPGKRK------KEMTKQKEAPEAKKQKIEGS 304

Qy        457 RDTLPSNMYQKNGLHHGKYAVKKSRRTDVEDLTPNPKKLLQIGNELRKLNKVISDLTPVS 516
                | | |::                  : :| ||          : :|   ||:|   :
Db        305 EPTTPFNLF------------------IGNLNPN--------KSVAELKVAISELFAKN 337

Qy        517 EL-PLTARPRSRKEKNKLAFRACRLKKKAQYEANKVKLWGLNTEYDNLLFVINSIKQE 573
              :|  :  |  : ::   : | :    :||  |   :|::|             | ||| |
Db        338 DLAAVDVRTGTNRKFGYVDFESAEDLEKA-LELTGLKVFG------------NEIKLE 382


RESULT 10
ADE57828
ID   ADE57828 standard; protein; 712 AA.
XX
AC   ADE57828;
XX
DT   29-JAN-2004  (first entry)
XX
DE   Rat Protein AAA41732, SEQ ID NO 3693.
XX
KW   Rat; pain; neuronal tissue; gene therapy; spinal segmental nerve injury;
KW   chronic constriction injury; CCI; spared nerve injury; SNI; Chung.
XX
OS   Rattus norvegicus.
XX
PN   WO2003016475-A2.
XX
PD   27-FEB-2003.
XX
PF   14-AUG-2002; 2002WO-US025765.
XX
PR   14-AUG-2001; 2001US-0312147P.
PR   01-NOV-2001; 2001US-0346382P.
PR   26-NOV-2001; 2001US-0333347P.
XX
PA   (GEHO ) GEN HOSPITAL CORP.
PA   (FARB ) BAYER AG.
XX
PI   Woolf C,  D'urso D,  Befort K,  Costigan M;
XX
DR   WPI; 2003-268312/26.
DR   GENBANK; AAA41732.
XX
PT   New composition comprising two or more isolated polypeptides, useful for
PT   preparing a medicament for treating pain in an animal.
XX
PS   Claim 1; Page; 1017pp; English.
XX
CC   The invention discloses a composition comprising two or more isolated rat
CC   or human polynucleotides or a polynucleotide which represents a fragment,
CC   derivative or allelic variation of the nucleic acid sequence. Also
CC   claimed are a vector comprising the novel polynucleotide, a host cell
CC   comprising the vector, a method for identifying a nucleotide sequence
CC   which is differentially regulated in an animal subjected to pain and a
CC   kit to perform the method, an array, a method for identifying an agent
CC   that increases or decreases the expression of the polynucleotide sequence
CC   that is differentially expressed in neuronal tissue of a first animal
CC   subjected to pain, a method for identifying a compound which regulates
CC   the expression of a polynucleotide sequence which is differentially
CC   expressed in an animal subjected to pain, a method for identifying a
CC   compound that regulates the activity of one or more of the
CC   polynucleotides, a method for producing a pharmaceutical composition, a
CC   method for identifying a compound or small molecule that regulates the
CC   activity in an animal of one or more of the polypeptides given in the
CC   specification, a method for identifying a compound useful in treating
CC   pain and a pharmaceutical composition comprising the one or more
CC   polypeptides or their antibodies. The polynucleotide or the compound that
CC   modulates its activity is useful for preparing a medicament for treating
CC   pain (e.g. spinal segmental nerve injury (Chung), chronic constriction
CC   injury (CCI) and spared nerve injury (SNI)) in an animal (e.g. gene
CC   therapy). The sequence presented is a rat protein (shown in Table 2 of
CC   the specification) which is differentially expressed during pain. Note:
CC   The sequence data for this patent did not form part of the printed
CC   specification, but was obtained in electronic form directly from WIPO at
CC   ftp.wipo.int/pub/published_pct_sequences.
XX
SQ   Sequence 712 AA;

  Query Match             6.0%;  Score 202;  DB 7;  Length 712;
  Best Local Similarity   22.4%;  Pred. No. 3.1e-06;
  Matches  107;  Conservative   60;  Mismatches  171;  Indels  140;  Gaps   18;

Qy        129 DEEVISKTPTLAQLNSEDSQSVSDSLYYPDSLFSVKQNPLPSSFPGKKITSRAAAPVCSS 188
              : : ::  |   : :||| :   |     |     ::  :|    ||| |:  |  |  |
Db         12 ESKKMAPPPKEVEEDSEDEEMSEDE----DDSSGEEEVVIPQK-KGKKATTTPAKKVVVS 66

Qy        189 KTLQAEVPLSDCVQKASKPPSSTQIMVKTNMYHNEKVNFHVECKDYVKKAKVKINPVQQS 248
              :| :| ||     :||:  |               |       |  |  |||   |
Db         67 QTKKAAVPTP--AKKAAVTPG-------------KKAAATPAKKAVTPAKVVPTP---- 106

Qy        249 RPLLSQIHTDAAKENTCYCGAVAKRQEKKGMEPLQGHATPALPFKETQELLLSPLPQEGP 308
                        |:     |:    |||       |||   |
Db        107 ----------GKKGAAQAKALVPTPGKKGA------VTPAKGAKN-------------- 135

Qy        309 GSLAAGESSSLSASTSVSDSSQKKEEHNYSLFVSDNLGEQPTKCSP------EEDEEDEE 362
              |  |  | |          | | : |:           | :| | :|      :|||||::
Db        136 GKNAKKEDSDEDEDEEDEDDSDEDEDEEDEFEPPVVKGVKPAKAAPAAPASEDEDEEDDD 195

Qy        363 DVDDEDHDEGFGSEHELSENEEE----------------------EEEEEDYEDDKDD 398
              | ||:| ||    | |  ::|||                        ||||| |||:|:
```

```
Db        196 DEDDDDDDE---EEEEEDDSEEEVMEITPAKGKKTPAKVVPVKAKSVAEEEEDDEDDEDE 252

Qy        399 DISDTFSEPGYENDSVEDLKEVTSISSRKRGKRRYFWEYSEQLTPSQQ--ERMLRPSEWN 456

Db        253 EEDE--DEEDEEDDEDEDEEEEEEPVKAAPGKRK------KEMTKQKEAPEAKKQKIEGS 304

Qy        457 RDTLPSNMYQKNGLHHGKYAVKKSRRTDVEDLTPNPKKLLQIGNELRKLNKVISDLTPVS 516

Db        305 EPTTPFNLF------------------IGNLNPN--------KSVAELKVAISELFAKN 337

Qy        517 EL-PLTARPRSRKEKNKLAFRACRLKKKAQYEANKVKLWGLNTEYDNLLFVINSIKQE 573

Db        338 DLAAVDVRTGTNRKFGYVDFESAEDLEKA-LELTGLKVFG-----------NEIKLE 382


RESULT 11
ADE57830
ID   ADE57830 standard; protein; 712 AA.
XX
AC   ADE57830;
XX
DT   29-JAN-2004  (first entry)
XX
DE   Rat Protein AAA41732, SEQ ID NO 3695.
XX
KW   Rat; pain; neuronal tissue; gene therapy; spinal segmental nerve injury;
KW   chronic constriction injury; CCI; spared nerve injury; SNI; Chung.
XX
OS   Rattus norvegicus.
XX
PN   WO2003016475-A2.
XX
PD   27-FEB-2003.
XX
PF   14-AUG-2002; 2002WO-US025765.
XX
PR   14-AUG-2001; 2001US-0312147P.
PR   01-NOV-2001; 2001US-0346382P.
PR   26-NOV-2001; 2001US-0333347P.
XX
PA   (GEHO ) GEN HOSPITAL CORP.
PA   (FARB ) BAYER AG.
XX
PI   Woolf C,  D'urso D,  Befort K,  Costigan M;
XX
DR   WPI; 2003-268312/26.
DR   GENBANK; AAA41732.
XX
PT   New composition comprising two or more isolated polypeptides, useful for
PT   preparing a medicament for treating pain in an animal.
XX
PS   Claim 1; Page; 1017pp; English.
XX
CC   The invention discloses a composition comprising two or more isolated rat
CC   or human polynucleotides or a polynucleotide which represents a fragment,
CC   derivative or allelic variation of the nucleic acid sequence. Also
CC   claimed are a vector comprising the novel polynucleotide, a host cell
CC   comprising the vector, a method for identifying a nucleotide sequence
CC   which is differentially regulated in an animal subjected to pain and a
CC   kit to perform the method, an array, a method for identifying an agent
CC   that increases or decreases the expression of the polynucleotide sequence
CC   that is differentially expressed in neuronal tissue of a first animal
CC   subjected to pain, a method for identifying a compound which regulates
CC   the expression of a polynucleotide sequence which is differentially
CC   expressed in an animal subjected to pain, a method for identifying a
CC   compound that regulates the activity of one or more of the
CC   polynucleotides, a method for producing a pharmaceutical composition, a
CC   method for identifying a compound or small molecule that regulates the
CC   activity in an animal of one or more of the polypeptides given in the
CC   specification, a method for identifying a compound useful in treating
CC   pain and a pharmaceutical composition comprising the one or more
CC   polypeptides or their antibodies. The polynucleotide or the compound that
CC   modulates its activity is useful for preparing a medicament for treating
CC   pain (e.g. spinal segmental nerve injury (Chung), chronic constriction
CC   injury (CCI) and spared nerve injury (SNI)) in an animal (e.g. gene
CC   therapy). The sequence presented is a rat protein (shown in Table 2 of
CC   the specification) which is differentially expressed during pain. Note:
CC   The sequence data for this patent did not form part of the printed
CC   specification, but was obtained in electronic form directly from WIPO at
CC   ftp.wipo.int/pub/published_pct_sequences.
XX
SQ   Sequence 712 AA;

  Query Match             6.0%;  Score 202;  DB 7;  Length 712;
  Best Local Similarity   22.4%;  Pred. No. 3.1e-06;
  Matches  107;  Conservative   60;  Mismatches  171;  Indels  140;  Gaps   18;

Qy        129 DEEVISKTPTLAQLNSEDSQSVSDSLYYPDSLFSVKQNPLPSSFPGKKITSRAAAPVCSS 188

Db         12 ESKKMAPPPKEVEEDSEDEEMSEDE----DDSSGEEEVVIPQK-KGKKATTTPAKKVVVS 66

Qy        189 KTLQAEVPLSDCVQKASKPPSSTQIMVKTNMYHNEKVNFHVECKDYVKKAKVKINPVQQS 248

Db         67 QTKKAAVPTP--AKKAAVTPG--------------KKAAATPAKKAVTPAKVVPTP---- 106

Qy        249 RPLLSQIHTDAAKENTCYCGAVAKRQEKKGMEPLQGHATPALPFKETQELLLSPLPQEGP 308

Db        107 -----------GKKGAAQAKALVPTPGKKGA------VTPAKGAKN-------------- 135

Qy        309 GSLAAGESSSLSASTSVSDSSQKKEEHNYSLFVSDNLGEQPTKCSP------EEDEEDEE 362

Db        136 GKNAKKEDSDEDEDEEDEDDSDEDEDEEDEFEPPVVKGVKPAKAAPAAPASEDEDEEDDD 195

Qy        363 DVDDEDHDEGFGSEHELSENEEE-----------------------EEEEEDYEDDKDD 398

Db        196 DEDDDDDDE---EEEEEDDSEEEVMEITPAKGKKTPAKVVPVKAKSVAEEEEDDEDDEDE 252

Qy        399 DISDTFSEPGYENDSVEDLKEVTSISSRKRGKRRYFWEYSEQLTPSQQ--ERMLRPSEWN 456

Db        253 EEDE--DEEDEEDDEDEDEEEEEEPVKAAPGKRK------KEMTKQKEAPEAKKQKIEGS 304

Qy        457 RDTLPSNMYQKNGLHHGKYAVKKSRRTDVEDLTPNPKKLLQIGNELRKLNKVISDLTPVS 516

Db        305 EPTTPFNLF------------------IGNLNPN--------KSVAELKVAISELFAKN 337

Qy        517 EL-PLTARPRSRKEKNKLAFRACRLKKKAQYEANKVKLWGLNTEYDNLLFVINSIKQE 573

Db        338 DLAAVDVRTGTNRKFGYVDFESAEDLEKA-LELTGLKVFG-----------NEIKLE 382


RESULT 12
ABG29275
ID   ABG29275 standard; protein; 160 AA.
XX
AC   ABG29275;
XX
DT   18-FEB-2002  (first entry)
XX
DE   Novel human diagnostic protein #29266.
XX
KW   Human; chromosome mapping; gene mapping; gene therapy; forensic;
KW   food supplement; medical imaging; diagnostic; genetic disorder.
XX
OS   Homo sapiens.
XX
PN   WO200175067-A2.
XX
PD   11-OCT-2001.
XX
PF   30-MAR-2001; 2001WO-US008631.
XX
PR   31-MAR-2000; 2000US-00540217.
PR   23-AUG-2000; 2000US-00649167.
XX
PA   (HYSE-) HYSEQ INC.
XX
```

```
PI   Drmanac RT,  Liu C,  Tang YT;
XX
DR   WPI; 2001-639362/73.
DR   N-PSDB; AAS93462.
XX
PT   New isolated polynucleotide and encoded polypeptides, useful in
PT   diagnostics, forensics, gene mapping, identification of mutations
PT   responsible for genetic disorders or other traits and to assess
PT   biodiversity.
XX
PS   Claim 20; SEQ ID NO 59634; 103pp; English.
XX
CC   The invention relates to isolated polynucleotide (I) and polypeptide (II)
CC   sequences. (I) is useful as hybridisation probes, polymerase chain
CC   reaction (PCR) primers, oligomers, and for chromosome and gene mapping,
CC   and in recombinant production of (II). The polynucleotides are also used
CC   in diagnostics as expressed sequence tags for identifying expressed
CC   genes. (I) is useful in gene therapy techniques to restore normal
CC   activity of (II) or to treat disease states involving (II). (II) is
CC   useful for generating antibodies against it, detecting or quantitating a
CC   polypeptide in tissue, as molecular weight markers and as a food
CC   supplement. (II) and its binding partners are useful in medical imaging
CC   of sites expressing (II). (I) and (II) are useful for treating disorders
CC   involving aberrant protein expression or biological activity. The
CC   polypeptide and polynucleotide sequences have applications in
CC   diagnostics, forensics, gene mapping, identification of mutations
CC   responsible for genetic disorders or other traits to assess biodiversity
CC   and to produce other types of data and products dependent on DNA and
CC   amino acid sequences. ABG00010-ABG30377 represent novel human diagnostic
CC   amino acid sequences of the invention. Note: The sequence data for this
CC   patent did not appear in the printed specification, but was obtained in
CC   electronic format directly from WIPO at
CC   ftp.wipo.int/pub/published_pct_sequences
XX
SQ   Sequence 160 AA;

  Query Match             6.0%;  Score 200.5;  DB 4;  Length 160;
  Best Local Similarity   73.0%;  Pred. No. 4.4e-07;
  Matches   46;  Conservative    2;  Mismatches   12;  Indels    3;  Gaps    2;

Qy        540 LKKKAQYEANKVKLW-GLNTEYDNLLFVINSIKQEIVNRVQNPRDERGPNMGQKLEILIK 598
              | |   |: :   ||     |  ||||||||||||||||||||||||||||||||||||
Db          8 LTKNHGYDHS--SLWRSTALEPYNLLFVINSIKQEIVNRVQNPRDERGPNMGQKLEILIK 65

Qy        599 DTL 601
              |||
Db         66 DTL 68


RESULT 13
AAY96255
ID   AAY96255 standard; protein; 1162 AA.
XX
AC   AAY96255;
XX
DT   12-SEP-2003  (revised)
DT   11-SEP-2000  (first entry)
XX
DE   Kaposi's sarcoma-associated herpesvirus LANA.
XX
KW   Kaposi's sarcoma-associated herpesvirus; KSHV; rhadino virus;
KW   latency-associated nuclear antigen; LANA; gamma-2 herpes virus;
KW   Human herpes virus 8; HHV8; rhadino virus cis-acting element; RVCAE;
KW   Kaposi's sarcoma; primary effusion lymphoma; PEL;
KW   human immunodeficiency virus; HIV; multicentric Castleman's disease.
XX
OS   Human herpesvirus 8.
XX
FH   Key             Location/Qualifiers
FT   Domain          14. .17
FT                   /note= "nuclear localisation signal, NLS"
FT   Domain          64. .70
FT                   /note= "nuclear localisation signal, NLS"
FT   Region          320. .429
FT                   /note= "acidic repeat region"
FT   Region          430. .549
FT                   /note= "Gln, Glu, Pro-rich region"
FT   Region          550. .589
FT                   /note= "Gln, Glu, Pro, Arg-rich region"
FT   Region          590. .759
FT                   /note= "Gln, Glu, Asp-rich region"
FT   Region          760. .840
FT                   /note= "Gln, Glu-rich region"
XX
PN   WO200029626-A1.
XX
PD   25-MAY-2000.
XX
PF   19-NOV-1999;   99WO-US027508.
XX
PR   19-NOV-1998;   98US-00109422.
PR   21-APR-1999;   99US-00298568.
XX
PA   (KIEF/) KIEFF E D.
PA   (BALL/) BALLESTAS M E.
PA   (KAYE/) KAYE K M.
XX
PI   Kieff ED,  Ballestas ME,  Kaye KM;
XX
DR   WPI; 2000-387829/33.
DR   N-PSDB; AAA30290.
XX
PT   Treating or preventing a disease associated with rhodino virus infection
PT   in a mammal which includes Kaposi's Sarcoma and Primary Effusion
PT   Lymphoma.
XX
PS   Disclosure; Fig 7; 70pp; English.
XX
CC   The present sequence is the Kaposi's sarcoma-associated herpesvirus,
CC   (KSHV) latency-associated nuclear antigen (LANA). KSHV is also known as
CC   Human Herpes Virus 8 (HHV8) and belongs to the rhadino virus, or gamma-2
CC   herpes virus class. The LANA protein is necessary for the efficient
CC   persistence of rhadino virus DNA in mammalian cells. Persistent rhadino
CC   virus infection is implicated in a variety of diseases e.g. Kaposi's
CC   Sarcoma (KS), Primary Effusion Lymphoma (PEL) and multicentric
CC   Castleman's disease. In addition, KS is a common malignancy in HIV
CC   patients. KSHV persists in host cells in a latent form. One of the few
CC   genes expressed from the latent viral DNA is LANA. LANA associates with
CC   both human chromosomes and with the rhadino virus cis-acting element
CC   (RVCAE), thereby providing a tethering function: the KSHV DNA episome is
CC   "tied" to the host chromosomes. This allows the viral DNA to persist in
CC   the host cell. The present sequence may be used to screen and identify
CC   molecules that inhibit LANA interaction with RVCAE, thereby interfering
CC   with the latency cycle of this virus. Potential antiviral treatments for
CC   the above mentioned diseases may therefore be based on LANA deregulation.
CC   (Updated on 12-SEP-2003 to standardise OS field)
XX
SQ   Sequence 1162 AA;

  Query Match             5.9%;  Score 196.5;  DB 3;  Length 1162;
  Best Local Similarity   24.2%;  Pred. No. 1.7e-05;
  Matches   96;  Conservative   52;  Mismatches  145;  Indels  103;  Gaps   14;

Qy        124 SSPYQDEEVISKTPTLAQLNSEDSQSVSDSLYYPDSLFSVKQNPLPSSFPGKKITSRAAA 183
              |||       :|   |        |: | | :|     |  | :: || ||  |:  :|
Db        111 SSPIPPSHPVSPGTT-------DTHSPSPALPPTQSPESSQRPPL-SSPTGRPDSSTPMR 162

Qy        184 PVCSSKTLQAEVPLSDCVQKASK------PPSSTQIMVKTNMYHNEKVNFHVECKDYVKK 237
              |  | :|     | :   :  ||       ||: : : |  :                 :
Db        163 PPPSQQTTPPHSPTTPPPEPPSKSSPDSLAPSTLRSLRKRRLSS-------------PQ 208

Qy        238 AKVKINPVQQSRPLL--------SQIHTDAAKENTCYCGAVAKRQEKKGMEPLQ----GH 285
                 :||: || |:           ::   | |:  | |: :          |      |
Db        209 GPSTLNPICQSPPVSPPRCDFANRSVYPPWATESPIYVGSSSDGDTPPRQPPTSPISIGS 268

Qy        286 ATPALPF--KETQELLL--------SPLPQEGPGSLAAGESSSLSASTSVSDSSQKKEEH 335
```

```
            ::|:      :|  |:|        |   :|   :  |||    :    :|  ||  ::
Db        269 SSPSEGSWGDDTAMLVLLAEIAEEASKNEKECSENNQAGED---NGDNEISKESQVDKDD 325

Qy        336 NYSLFVSDNLGEQPTKCSPEEDEEDEE--------------------------DVDDED 368
              |      ||  ::  : : ||||||:|                          : |||:
Db        326 N------DNKDDEEEQETDEEDEEDDEEDDEEDDEEDDEEDDEEDDEEDDEEDDEEDEE 379

Qy        369 HDEGFGSEHELSENEEEEEEEEDYEDDKDDDISDTFSEPGYENDSVEDLKE--------- 419
               ||    | :  }:|||:||||| ||| |:|  |       |:|  || ||
Db        380 DDEEDDEEEDEEEDEEEDEEEEDEEDDDDEDNEDE------EDDEEEDKKEDEEDGGDGN 433

Qy        420 -VTSISSRKRGKRRYFWEYSEQLTPSQQERMLRPSE 454
                 || | :: :     |  :|  | ||| :  | :
Db        434 KTLSIQSSQQQQEPQQQE-PQQQEPQQQEPLQEPQQ 468


RESULT 14
AAY58500
ID   AAY58500 standard; protein; 1162 AA.
XX
AC   AAY58500;
XX
DT   06-AUG-2003  (revised)
DT   10-APR-2000  (first entry)
XX
DE   HHV8 ORF 73 protein, SEQ ID NO:21.
XX
KW   HHV8; detection; diagnosis; Kaposi's sarcoma; AIDS; immunogen; antigen.
XX
OS   Human herpesvirus 8.
XX
FH   Key             Location/Qualifiers
FT   Misc-difference 96
FT                   /label= unknown
XX
PN   WO9961909-A2.
XX
PD   02-DEC-1999.
XX
PF   26-MAY-1999;   99WO-US011407.
XX
PR   26-MAY-1998;   98US-0086695P.
XX
PA   (USSH ) US DEPT HEALTH & HUMAN SERVICES.
XX
PI   Pau C;
XX
DR   WPI; 2000-097142/08.
XX
PT   New methods and compositions for the detection of human herpesvirus.
XX
PS   Claim 2; Page 59-62; 68pp; English.
XX
CC   Sequences AAY58480-Y58532 represent immunogenic polypeptides derived from
CC   human herpes virus type 8 (HHV8, a gammaherpesvirus). HHV8 plays an
CC   important role in the pathogenesis of AIDS-related Kaposi's sarcoma. The
CC   invention relates to a novel method of detecting the presence of human
CC   herpesvirus 8 in a biological sample using peptides representative of
CC   dominant antigenic regions of HHV8. The method comprises contacting one
CC   or more isolated, immunogenic HHV8 peptides with an antibody-containing
CC   biological sample, and detecting the formation of a complex between the
CC   peptide and the antibody. The presence of a peptide-antibody complex
CC   indicates the presence of human herpesvirus 8. The detection of HHV8
CC   infection can be used to diagnose AIDS-associated Kaposi's sarcoma. The
CC   HHV8-specific antibodies are useful therapeutically when for the passive
CC   immunisation of a human against HHV8 infection, thereby reducing HHV8
CC   related disease. The detection assays are highly specific, sensitive and
CC   accurate. Early detection and treatment of Kaposi's sarcoma could
CC   diminish the severity of symptoms related to AIDS and the sensitive
CC   techniques could reduce erroneous characterisations of skin disorders.
CC   Previous assays for HHV8 antibodies such as immunofluorescence assays,
CC   immunoblots and enzyme immunosassays lack the sensitivity and accuracy
CC   needed for reliable diagnosis of Kaposi's sarcoma. Further advantages of
CC   the assays are that reproducible results are obtained and the method is
CC   suitable for rapid throughput and screening of samples economically.
CC   (Updated on 06-AUG-2003 to correct OS field.)
XX
SQ   Sequence 1162 AA;

  Query Match             5.9%;  Score 196.5;  DB 3;  Length 1162;
  Best Local Similarity   24.2%;  Pred. No. 1.7e-05;
  Matches   96;  Conservative   52;  Mismatches  145;  Indels  103;  Gaps   14;

Qy        124 SSPYQDEEVISKTPTLAQLNSEDSQSVSDSLYYPDSLFSVKQNPLPSSFPGKKITSRAAA 183
              |||      :|    |        |: | | :|    |  | :: || ||  |:  :|
Db        111 SSPIPPSHPVSPGTT-------DTHSPSPALPPTQSPESSQRPPL-SSPTGRPDSSTPMR 162

Qy        184 PVCSSKTLQAEVPLSDCVQKASK------PPSSTQIMVKTNMYHNEKVNFHVECKDYVKK 237
              |  | :|     | :   :  ||       ||: : : |  :                 :
Db        163 PPPSQQTTPPHSPTTPPPEPPSKSSPDSLAPSTLRSLRKRRLSS--------------PQ 208

Qy        238 AKVKINPVQQSRPLL--------SQIHTDAAKENTCYCGAVAKRQEKKGMEPLQ----GH 285
                :||: || |:           ::   | |:  | |: :          |      |
Db        209 GPSTLNPICQSPPVSPPRCDFANRSVYPPWATESPIYVGSSSDGDTPPRQPPTSPISIGS 268

Qy        286 ATPALPF--KETQELLL--------SPLPQEGPGSLAAGESSSLSASTSVSDSSQKKEEH 335
              ::|:      :|  |:|        |  :|   :  |||    :    :|  ||  ::
Db        269 SSPSEGSWGDDTAMLVLLAEIAEEASKNEKECSENNQAGED---NGDNEISKESQVDKDD 325

Qy        336 NYSLFVSDNLGEQPTKCSPEEDEEDEE--------------------------DVDDED 368
              |      ||  ::  : : ||||||:|                          : |||:
Db        326 N------DNKDDEEEQETDEEDEEDDEEDDEEDDEEDDEEDDEEDDEEDDEEDDEEDEE 379

Qy        369 HDEGFGSEHELSENEEEEEEEEDYEDDKDDDISDTFSEPGYENDSVEDLKE--------- 419
               ||    | :  |:|||:||||| ||| |:|  |       |:|  || ||
Db        380 DDEEDDEEEDEEEDEEEDEEEEDEEDDDDEDNEDE------EDDEEEDKKEDEEDGGDGN 433

Qy        420 -VTSISSRKRGKRRYFWEYSEQLTPSQQERMLRPSE 454
                 || | :: :     |  :|  | ||| :  | :
Db        434 KTLSIQSSQQQQEPQQQE-PQQQEPQQQEPLQEPQQ 468


RESULT 15
AAB62331
ID   AAB62331 standard; protein; 1162 AA.
XX
AC   AAB62331;
XX
DT   06-AUG-2003  (revised)
DT   29-JUN-2001  (first entry)
XX
DE   Amino acid sequence of KSHV tethering protein LANA.
XX
KW   Histone H1; tethering protein; LANA; gene therapy; multiple sclerosis;
KW   Parkinson's disease; Huntington disease; diabetes; human herpesvirus 8;
KW   KSHV; latency-associated nuclear antigen; LANA.
XX
OS   Human herpesvirus 8.
XX
PN   WO200125484-A2.
XX
PD   12-APR-2001.
XX
PF   29-SEP-2000; 2000WO-US026908.
XX
PR   01-OCT-1999;   99US-00410399.
XX
PA   (UNMI ) UNIV MICHIGAN.
XX
PI   Robertson ES,  Cotter MA;
XX
DR   WPI; 2001-281736/29.
DR   N-PSDB; AAF82901.
XX
PT   A composition for use in gene therapy comprises an expression vector that
PT   includes a nucleic acid sequence encoding a nucleic acid binding protein.
```

XX
PS Disclosure; Fig 9B; 60pp; English.
XX
CC The invention provides a composition comprising nucleic acid, histone H1
CC protein and expression vector operationally encoding a protein suitable
CC for tethering the nucleic acid to the histone H1 protein, where the
CC tethering protein is LANA. The composition is useful in aiding the
CC retention of the viral DNA in the host cell. The viral vector encodes a
CC protein suitable for tethering DNA to Histone H1. Methods for screening
CC for compounds which are agonistic or antagonistic for the tethering of
CC viral proteins to histone H1 and DNA binding sites are useful for
CC developing the method of viral transfer. The composition has applications
CC to gene therapy, including the treatment of multiple sclerosis,
CC Parkinson's disease, Huntington disease and diabetes. The present
CC sequence represents the amino acid sequence of the Kaposi's sarcoma
CC associated herpesvirus (human herpesvirus 8) latency-associated nuclear
CC antigen (LANA), which acts as a tethering protein. (Updated on 06-AUG-
CC 2003 to correct OS field.)
XX
SQ Sequence 1162 AA;

Query Match 5.9%; Score 196.5; DB 4; Length 1162;
Best Local Similarity 24.2%; Pred. No. 1.7e-05;
Matches 96; Conservative 52; Mismatches 145; Indels 103; Gaps 14;

```
Qy      124 SSPYQDEEVISKTPTLAQLNSEDSQSVSDSLYYPDSLFSVKQNPLPSSFPGKKITSRAAA 183
            |||      :|   |       |: | | :|    |  | :: || ||  |:  :|
Db      111 SSPIPPSHPVSPGTT-------DTHSPSPALPPTQSPESSQRPPL-SSPTGRPDSSTPMR 162

Qy      184 PVCSSKTLQAEVPLSDCVQKASK------PPSSTQIMVKTNMYHNEKVNFHVECKDYVKK 237
            |  | :|     | :   :  ||       ||: : : |  :                 :
Db      163 PPPSQQTTPPHSPTTPPPEPPSKSSPDSLAPSTLRSLRKRRLSS-------------PQ 208

Qy      238 AKVKINPVQQSRPLL--------SQIHTDAAKENTCYCGAVAKRQEKKGMEPLQ----GH 285
              :||: || |:          ::   | |:  | |: :          |       |
Db      209 GPSTLNPICQSPPVSPPRCDFANRSVYPPWATESPIYVGSSSDGDTPPRQPPTSPISIGS 268

Qy      286 ATPALPF--KETQELLL--------SPLPQEGPGSLAAGESSSLSASTSVSDSSQKKEEH 335
            ::|:      :|  |:|        |   :|   :  |||    :    :|  ||  ::
Db      269 SSPSEGSWGDDTAMLVLLAEIAEEASKNEKECSENNQAGED---NGDNEISKESQVDKDD 325

Qy      336 NYSLFVSDNLGEQPTKCSPEEDEEDEE--------------------------DVDDED 368
            |      ||  ::  : : ||||||:|                          : |||:
Db      326 N------DNKDDEEEQETDEEDEEDDEEDDEEDDEEDDEEDDEEDDEEDDEEDDEE 379

Qy      369 HDEGFGSEHELSENEEEEEEEEDYEDDKDDDISDTFSEPGYENDSVEDLKE--------- 419
             ||    | :  |:|||:||||| ||| |:|  |       |:|  || ||
Db      380 DDEEDDEEEDEEEDEEEDEEEEDEDDDDEDNEDE------EDDEEEDKKEDEEDGGDGN 433

Qy      420 -VTSISSRKRGKRRYFWEYSEQLTPSQQERMLRPSE 454
             || | :: :     |  :|  | ||| :  | :
Db      434 KTLSIQSSQQQQEPQQQE-PQQQEPQQQEPLQEPQQ 468
```

Search completed: February 28, 2006, 08:45:04
Job time : 222.872 secs

GenCore version 5.1.7

OM protein - protein search, using sw model

Run on: February 28, 2006, 08:53:26 ; Search time 54.985 Seconds
(without alignments)
960.804 Million cell updates/sec

Title: US-10-717-665A-44
Perfect score: 3349
Sequence: 1 MPQPSVSGMDPPFGDAFRSH.....TAEGNPTGGLVGLRIPTSKV 639

Scoring table: BLOSUM62
Gapop 10.0 , Gapext 0.5

Searched: 572060 seqs, 82675679 residues

Total number of hits satisfying chosen parameters: 572060

Minimum DB seq length: 0
Maximum DB seq length: 2000000000

Post-processing: Minimum Match 0%
Maximum Match 100%
Listing first 45 summaries

Database : Issued_Patents_AA:*
1: /cgn2_6/ptodata/1/iaa/5_COMB.pep:*
2: /cgn2_6/ptodata/1/iaa/6_COMB.pep:*
3: /cgn2_6/ptodata/1/iaa/H_COMB.pep:*
4: /cgn2_6/ptodata/1/iaa/PCTUS_COMB.pep:*
5: /cgn2_6/ptodata/1/iaa/RE_COMB.pep:*
6: /cgn2_6/ptodata/1/iaa/backfiles1.pep:*

Pred. No. is the number of results predicted by chance to have a score greater than or equal to the score of the result being printed, and is derived by analysis of the total score distribution.

SUMMARIES

Result No.	Score	% Query Match	Length	DB	ID	Description
1	3349	100.0	639	2	US-10-164-595-44	Sequence 44, Appl
2	434	13.0	91	2	US-10-164-595-75	Sequence 75, Appl
3	324	9.7	141	2	US-09-270-767-41799	Sequence 41799, A
4	324	9.7	141	2	US-09-270-767-57041	Sequence 57041, A
5	196.5	5.9	1162	1	US-08-728-323A-2	Sequence 2, Appli
6	196.5	5.9	1162	2	US-09-298-568-2	Sequence 2, Appli
7	196.5	5.9	1162	2	US-09-410-399-2	Sequence 2, Appli
8	196.5	5.9	1162	2	US-09-894-273-2	Sequence 2, Appli
9	184	5.5	764	1	US-08-375-300-4	Sequence 4, Appli
10	184	5.5	764	2	US-09-177-431-4	Sequence 4, Appli
11	184	5.5	764	4	PCT-US95-16930-4	Sequence 4, Appli
12	184	5.5	1089	1	US-08-375-300-2	Sequence 2, Appli
13	184	5.5	1089	2	US-09-177-431-2	Sequence 2, Appli
14	184	5.5	1089	4	PCT-US95-16930-2	Sequence 2, Appli
15	176.5	5.3	428	2	US-09-248-796A-15197	Sequence 15197, A
16	174	5.2	706	2	US-09-538-092-957	Sequence 957, App
17	174	5.2	747	2	US-09-949-016-10040	Sequence 10040, A
18	170.5	5.1	311	2	US-09-902-540-10544	Sequence 10544, A
19	170.5	5.1	1231	2	US-08-714-741-41	Sequence 41, Appl
20	166.5	5.0	412	1	US-08-741-134-2	Sequence 2, Appli
21	166.5	5.0	542	2	US-08-935-855-22	Sequence 22, Appl
22	166.5	5.0	709	2	US-09-248-796A-19045	Sequence 19045, A
23	164.5	4.9	687	2	US-10-104-047-2651	Sequence 2651, Ap
24	164.5	4.9	1102	2	US-09-949-016-8427	Sequence 8427, Ap
25	163.5	4.9	1972	2	US-09-418-710-21	Sequence 21, Appl
26	163.5	4.9	1972	2	US-09-839-479-21	Sequence 21, Appl
27	162.5	4.9	714	1	US-08-990-114-3	Sequence 3, Appli
28	162.5	4.9	714	2	US-09-241-333-3	Sequence 3, Appli
29	161.5	4.8	982	2	US-09-248-796A-20628	Sequence 20628, A
30	161	4.8	831	1	US-09-047-026A-4	Sequence 4, Appli
31	160.5	4.8	797	2	US-09-949-016-9676	Sequence 9676, Ap
32	160.5	4.8	1969	2	US-09-418-710-72	Sequence 72, Appl
33	160.5	4.8	1969	2	US-09-839-479-71	Sequence 71, Appl
34	160	4.8	392	1	US-08-822-701-2	Sequence 2, Appli
35	160	4.8	392	2	US-08-935-855-2	Sequence 2, Appli
36	160	4.8	868	2	US-09-248-796A-16660	Sequence 16660, A
37	159	4.7	785	2	US-09-248-796A-24492	Sequence 24492, A
38	159	4.7	1085	1	US-08-431-080-28	Sequence 28, Appl
39	159	4.7	1085	1	US-08-938-534-28	Sequence 28, Appl
40	159	4.7	1085	2	US-09-345-294-28	Sequence 28, Appl
41	158	4.7	546	2	US-08-935-855-20	Sequence 20, Appl
42	158	4.7	546	2	US-09-538-092-827	Sequence 827, App
43	158	4.7	546	2	US-09-949-016-6265	Sequence 6265, Ap
44	158	4.7	553	2	US-09-949-016-7397	Sequence 7397, Ap
45	158	4.7	1089	2	US-09-949-016-10326	Sequence 10326, A

ALIGNMENTS

RESULT 1
US-10-164-595-44
; Sequence 44, Application US/10164595
; Patent No. 6657054
; GENERAL INFORMATION:
; APPLICANT: OriGene Technologies, Inc
; TITLE OF INVENTION: Regulated Angiogenesis Genes and Polypeptides
; FILE REFERENCE: 1U 103 R1
; CURRENT APPLICATION NUMBER: US/10/164,595
; CURRENT FILING DATE: 2002-06-10
; NUMBER OF SEQ ID NOS: 80
; SOFTWARE: PatentIn version 3.1
; SEQ ID NO 44
; LENGTH: 639
; TYPE: PRT
; ORGANISM: Homo sapiens
US-10-164-595-44

Query Match 100.0%; Score 3349; DB 2; Length 639;
Best Local Similarity 100.0%; Pred. No. 3.7e-279;
Matches 639; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

```
Qy          1 MPQPSVSGMDPPFGDAFRSHTFSEQTLMSTDLLANSSDPDFMYELDREMNYQQNPRDNFL 60
              ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db          1 MPQPSVSGMDPPFGDAFRSHTFSEQTLMSTDLLANSSDPDFMYELDREMNYQQNPRDNFL 60

Qy         61 SLEDCKDIENLESFTDVLDNEGALTSNWEQWDTYCEDLTKYTKLTSCDIWGTKEVDYLGL 120
              ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db         61 SLEDCKDIENLESFTDVLDNEGALTSNWEQWDTYCEDLTKYTKLTSCDIWGTKEVDYLGL 120

Qy        121 DDFSSPYQDEEVISKTPTLAQLNSEDSQSVSDSLYYPDSLFSVKQNPLPSSFPGKKITSR 180
              ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db        121 DDFSSPYQDEEVISKTPTLAQLNSEDSQSVSDSLYYPDSLFSVKQNPLPSSFPGKKITSR 180

Qy        181 AAAPVCSSKTLQAEVPLSDCVQKASKPPSSTQIMVKTNMYHNEKVNFHVECKDYVKKAKV 240
              ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db        181 AAAPVCSSKTLQAEVPLSDCVQKASKPPSSTQIMVKTNMYHNEKVNFHVECKDYVKKAKV 240

Qy        241 KINPVQQSRPLLSQIHTDAAKENTCYCGAVAKRQEKKGMEPLQGHATPALPFKETQELLL 300
              ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db        241 KINPVQQSRPLLSQIHTDAAKENTCYCGAVAKRQEKKGMEPLQGHATPALPFKETQELLL 300

Qy        301 SPLPQEGPGSLAAGESSSLSASTSVSDSSQKKEEHNYSLFVSDNLGEQPTKCSPEEDEED 360
              ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db        301 SPLPQEGPGSLAAGESSSLSASTSVSDSSQKKEEHNYSLFVSDNLGEQPTKCSPEEDEED 360

Qy        361 EEDVDDEDHDEGFGSEHELSENEEEEEEEEDYEDDKDDDISDTFSEPGYENDSVEDLKEV 420
              ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db        361 EEDVDDEDHDEGFGSEHELSENEEEEEEEEDYEDDKDDDISDTFSEPGYENDSVEDLKEV 420
```

```
Qy        421 TSISSRKRGKRRYFWEYSEQLTPSQQERMLRPSEWNRDTLPSNMYQKNGLHHGKYAVKKS 480
              ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db        421 TSISSRKRGKRRYFWEYSEQLTPSQQERMLRPSEWNRDTLPSNMYQKNGLHHGKYAVKKS 480

Qy        481 RRTDVEDLTPNPKKLLQIGNELRKLNKVISDLTPVSELPLTARPRSRKEKNKLAFRACRL 540
              ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db        481 RRTDVEDLTPNPKKLLQIGNELRKLNKVISDLTPVSELPLTARPRSRKEKNKLAFRACRL 540

Qy        541 KKKAQYEANKVKLWGLNTEYDNLLFVINSIKQEIVNRVQNPRDERGPNMGQKLEILIKDT 600
              ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db        541 KKKAQYEANKVKLWGLNTEYDNLLFVINSIKQEIVNRVQNPRDERGPNMGQKLEILIKDT 600

Qy        601 LGLPVAGQTSEFVNQVLEKTAEGNPTGGLVGLRIPTSKV 639
              ||||||||||||||||||||||||||||||||||||||||
Db        601 LGLPVAGQTSEFVNQVLEKTAEGNPTGGLVGLRIPTSKV 639
```

RESULT 2
US-10-164-595-75
; Sequence 75, Application US/10164595
; Patent No. 6657054
; GENERAL INFORMATION:
; APPLICANT: OriGene Technologies, Inc
; TITLE OF INVENTION: Regulated Angiogenesis Genes and Polypeptides
; FILE REFERENCE: 1U 103 R1
; CURRENT APPLICATION NUMBER: US/10/164,595
; CURRENT FILING DATE: 2002-06-10
; NUMBER OF SEQ ID NOS: 80
; SOFTWARE: PatentIn version 3.1
; SEQ ID NO 75
; LENGTH: 91
; TYPE: PRT
; ORGANISM: Homo sapiens
US-10-164-595-75

Query Match 13.0%; Score 434; DB 2; Length 91;
Best Local Similarity 100.0%; Pred. No. 9.3e-30;
Matches 83; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

```
Qy        449 MLRPSEWNRDTLPSNMYQKNGLHHGKYAVKKSRRTDVEDLTPNPKKLLQIGNELRKLNKV 508
              ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db          1 MLRPSEWNRDTLPSNMYQKNGLHHGKYAVKKSRRTDVEDLTPNPKKLLQIGNELRKLNKV 60

Qy        509 ISDLTPVSELPLTARPRSRKEKN 531
              |||||||||||||||||||||||
Db         61 ISDLTPVSELPLTARPRSRKEKN 83
```

RESULT 3
US-09-270-767-41799
; Sequence 41799, Application US/09270767
; Patent No. 6703491
; GENERAL INFORMATION:
; APPLICANT: Homburger et al.
; TITLE OF INVENTION: Nucleic acids and proteins of Drosophila melanogaster
; FILE REFERENCE: File Reference: 7326-094
; CURRENT APPLICATION NUMBER: US/09/270,767
; CURRENT FILING DATE: 1999-03-17
; NUMBER OF SEQ ID NOS: 62517
; SOFTWARE: PatentIn Ver. 2.0
; SEQ ID NO 41799
; LENGTH: 141
; TYPE: PRT
; ORGANISM: Drosophila melanogaster
US-09-270-767-41799

Query Match 9.7%; Score 324; DB 2; Length 141;
Best Local Similarity 49.3%; Pred. No. 5.2e-20;
Matches 71; Conservative 22; Mismatches 37; Indels 14; Gaps 2;

```
Qy        498 IGNELRKLNKVISDLTPVSELPLTARPRSRKEKNKLAFRACRLKKKAQYEANKVKLWGLN 557
              || || ||:: |:|:|||||||   ||:|||||||| |||||||||:||||:||:||
Qb          1 IGKELDKLSRTINDMTPVSELPFNVRPKSRKEKNKLASRACRLKKKAQHEANKIKLFGLE 60

Qy        558 TEYDNLLFVINSIKQEIVNRVQNPRDERGPNMGQKLE--------ILIKDTLGLPVAGQT 609
              |:  |:  |  :|| :| :         |  |:|:  |        |    : |: :|| :
Db         61 IEHKRLMNGIAELKQALVVK------HRTKNLGESTEEVDQQIARIYATASSGIRIAGGS 114

Qy        610 SEFVNQVLEKTAEGNPTGGLVGLR 633
              ::|||:|||    | | |||  ||
Db        115 TDFVNKVLENMRGGMPNGGLEELR 138
```

RESULT 4
US-09-270-767-57041
; Sequence 57041, Application US/09270767
; Patent No. 6703491
; GENERAL INFORMATION:
; APPLICANT: Homburger et al.
; TITLE OF INVENTION: Nucleic acids and proteins of Drosophila melanogaster
; FILE REFERENCE: File Reference: 7326-094
; CURRENT APPLICATION NUMBER: US/09/270,767
; CURRENT FILING DATE: 1999-03-17
; NUMBER OF SEQ ID NOS: 62517
; SOFTWARE: PatentIn Ver. 2.0
; SEQ ID NO 57041
; LENGTH: 141
; TYPE: PRT
; ORGANISM: Drosophila melanogaster
US-09-270-767-57041

Query Match 9.7%; Score 324; DB 2; Length 141;
Best Local Similarity 49.3%; Pred. No. 5.2e-20;
Matches 71; Conservative 22; Mismatches 37; Indels 14; Gaps 2;

```
Qy        498 IGNELRKLNKVISDLTPVSELPLTARPRSRKEKNKLAFRACRLKKKAQYEANKVKLWGLN 557
              || || ||:: |:|:|||||||   ||:|||||||| |||||||||:||||:||:||
Db          1 IGKELDKLSRTINDMTPVSELPFNVRPKSRKEKNKLASRACRLKKKAQHEANKIKLFGLE 60

Qy        558 TEYDNLLFVINSIKQEIVNRVQNPRDERGPNMGQKLE--------ILIKDTLGLPVAGQT 609
              |:  |:  |  :|| :| :         |  |:|:  |        |    : |: :|| :
Db         61 IEHKRLMNGIAELKQALVVK------HRTKNLGESTEEVDQQIARIYATASSGIRIAGGS 114

Qy        610 SEFVNQVLEKTAEGNPTGGLVGLR 633
              ::|||:|||    | | |||  ||
Db        115 TDFVNKVLENMRGGMPNGGLEELR 138
```

RESULT 5
US-08-728-323A-2
; Sequence 2, Application US/08728323A
; Patent No. 5948676
; GENERAL INFORMATION:
; APPLICANT: Chang, Yuan
; APPLICANT: Bohenzky, Roy A.
; APPLICANT: Russo, James J.
; APPLICANT: Edelman, Isidore S.
; APPLICANT: Moore, Patrick S.
; TITLE OF INVENTION: Immediate Early Protein From Kaposi's
; TITLE OF INVENTION: Sarcoma-Associated Herpesvirus, DNA
; TITLE OF INVENTION: Encoding Same And Uses Thereof
; NUMBER OF SEQUENCES: 21
; CORRESPONDENCE ADDRESS:
; ADDRESSEE: Cooper & Dunham LLP
; STREET: 1185 Avenue of the Americas
; CITY: New York
; STATE: New York
; COUNTRY: U.S.A.
; ZIP: 10036
; COMPUTER READABLE FORM:
; MEDIUM TYPE: Floppy disk
; COMPUTER: IBM PC compatible
; OPERATING SYSTEM: PC-DOS/MS-DOS
; SOFTWARE: PatentIn Release #1.0, Version #1.30
; CURRENT APPLICATION DATA:
; APPLICATION NUMBER: US/08/728,323A

```
;      FILING DATE:
;      CLASSIFICATION:  435
;    ATTORNEY/AGENT INFORMATION:
;      NAME:  White, John P.
;      REGISTRATION NUMBER:  28,678
;      REFERENCE/DOCKET NUMBER:  0575/52268/JPW/MSC/SKS
;    TELECOMMUNICATION INFORMATION:
;      TELEPHONE:  212-278-0400
;      TELEFAX:  212-391-0525
;  INFORMATION FOR SEQ ID NO:  2:
;    SEQUENCE CHARACTERISTICS:
;     LENGTH: 1162 amino acids
;      TYPE:  amino acid
;      TOPOLOGY:  linear
;    MOLECULE TYPE:  protein
US-08-728-323A-2
```

Query Match 5.9%; Score 196.5; DB 1; Length 1162;
Best Local Similarity 24.2%; Pred. No. 1.2e-07;
Matches 96; Conservative 52; Mismatches 145; Indels 103; Gaps 14;

```
Qy        124 SSPYQDEEVISKTPTLAQLNSEDSQSVSDSLYYPDSLFSVKQNPLPSSFPGKKITSRAAA 183
              |||       :|   |        |: | | :|     |  | :: || ||  |:  :|
Db        111 SSPIPPSHPVSPGTT-------DTHSPSPALPPTQSPESSQRPPL-SSPTGRPDSSTPMR 162

Qy        184 PVCSSKTLQAEVPLSDCVQKASK------PPSSTQIMVKTNMYHNEKVNFHVECKDYVKK 237
              |  | :|     | :   :  ||       ||: : : |  :                 :
Db        163 PPPSQQTTPPHSPTTPPPEPPSKSSPDSLAPSTLRSLRKRRLSS--------------PQ 208

Qy        238 AKVKINPVQQSRPLL--------SQIHTDAAKENTCYCGAVAKRQEKKGMEPLQ----GH 285
                :||: || |:           ::   | |:  | |: :          |      |
Db        209 GPSTLNPICQSPPVSPPRCDFANRSVYPPWATESPIYVGSSSDGDTPPRQPPTSPISIGS 268

Qy        286 ATPALPF--KETQELLL--------SPLPQEGPGSLAAGESSSLSASTSVSDSSQKKEEH 335
              ::|:      :|  |:|        |   :|   :  |||    :   :|  ||  ::
Db        269 SSPSEGSWGDDTAMLVLLAEIAEEASKNEKECSENNQAGED---NGDNEISKESQVDKDD 325

Qy        336 NYSLFVSDNLGEQPTKCSPEEDEEDEE--------------------------DVDDED 368
              |      ||  ::  : : ||||||:|                          : |||:
Db        326 N------DNKDDEEEQETDEEDEEDDEEDDEEDDEEDDEEDDEEDDEEDDEEDDEEDEE 379

Qy        369 HDEGFGSEHELSENEEEEEEEEDYEDDKDDDISDTFSEPGYENDSVEDLKE--------- 419
               ||    | :  |:|||:|||| ||| |:|  |       |:|  || ||
Db        380 DDEEDDEEEDEEEDEEEDEEEEDEDDDDEDNEDE------EDDEEEDKKEDEEDGGDGN 433

Qy        420 -VTSISSRKRGKRRYFWEYSEQLTPSQQERMLRPSE 454
               || | :: :    |  :|  | ||| :  | :
Db        434 KTLSIQSSQQQQEPQQQE-PQQQEPQQQEPLQEPQQ 468
```

```
RESULT 6
US-09-298-568-2
; Sequence 2, Application US/09298568
; Patent No. 6322792
; GENERAL INFORMATION:
;  APPLICANT: Kieff, Elliott D.
;  APPLICANT: Ballestas, Mary E.
;  APPLICANT: Kaye, Kenneth M.
;  TITLE OF INVENTION: RHADINO VIRUS LANA ACTS IN TRANS ON A UNIT OF RHADINO
;  TITLE OF INVENTION: VIRUS DNA TO MEDIATE EFFICIENT EPISOME PERSISTENCE
;  FILE REFERENCE: 16412-10001R
;  CURRENT APPLICATION NUMBER: US/09/298,568
;  CURRENT FILING DATE: 1999-04-21
;  EARLIER APPLICATION NUMBER: US 60/109,422
;  EARLIER FILING DATE: 1998-11-19
;  NUMBER OF SEQ ID NOS: 3
;  SOFTWARE: PatentIn Ver. 2.0
; SEQ ID NO 2
;   LENGTH: 1162
;   TYPE: PRT
;   ORGANISM: Kaposi's sarcoma-associated herpesvirus
US-09-298-568-2
```

Query Match 5.9%; Score 196.5; DB 2; Length 1162;
Best Local Similarity 24.2%; Pred. No. 1.2e-07;
Matches 96; Conservative 52; Mismatches 145; Indels 103; Gaps 14;

```
Qy        124 SSPYQDEEVISKTPTLAQLNSEDSQSVSDSLYYPDSLFSVKQNPLPSSFPGKKITSRAAA 183
              |||       :|   |        |: | | :|     |  | :: || ||  |:  :|
Db        111 SSPIPPSHPVSPGTT-------DTHSPSPALPPTQSPESSQRPPL-SSPTGRPDSSTPMR 162

Qy        184 PVCSSKTLQAEVPLSDCVQKASK------PPSSTQIMVKTNMYHNEKVNFHVECKDYVKK 237
              |  | :|     | :   :  ||       ||: : : |  :                 :
Db        163 PPPSQQTTPPHSPTTPPPEPPSKSSPDSLAPSTLRSLRKRRLSS--------------PQ 208

Qy        238 AKVKINPVQQSRPLL--------SQIHTDAAKENTCYCGAVAKRQEKKGMEPLQ----GH 285
                :||: || |:           ::   | |:  | |: :          |      |
Db        209 GPSTLNPICQSPPVSPPRCDFANRSVYPPWATESPIYVGSSSDGDTPPRQPPTSPISIGS 268

Qy        286 ATPALPF--KETQELLL--------SPLPQEGPGSLAAGESSSLSASTSVSDSSQKKEEH 335
              ::|:      :|  |:|        |   :|   :  |||    :   :|  ||  ::
Db        269 SSPSEGSWGDDTAMLVLLAEIAEEASKNEKECSENNQAGED---NGDNEISKESQVDKDD 325

Qy        336 NYSLFVSDNLGEQPTKCSPEEDEEDEE--------------------------DVDDED 368
              |      ||  ::  : : ||||||:|                          : |||:
Db        326 N------DNKDDEEEQETDEEDEEDDEEDDEEDDEEDDEEDDEEDDEEDDEEDDEEDEE 379

Qy        369 HDEGFGSEHELSENEEEEEEEEDYEDDKDDDISDTFSEPGYENDSVEDLKE--------- 419
               ||    | :  |:|||:|||| ||| |:|  |       |:|  || ||
Db        380 DDEEDDEEEDEEEDEEEDEEEEDEDDDDEDNEDE------EDDEEEDKKEDEEDGGDGN 433

Qy        420 -VTSISSRKRGKRRYFWEYSEQLTPSQQERMLRPSE 454
               || | :: :    |  :|  | ||| :  | :
Db        434 KTLSIQSSQQQQEPQQQE-PQQQEPQQQEPLQEPQQ 468
```

```
RESULT 7
US-09-410-399-2
; Sequence 2, Application US/09410399
; Patent No. 6482587
; GENERAL INFORMATION:
;  APPLICANT:  Robertson, Erle S.
;  APPLICANT:  Cotter, Murray A.
;  TITLE OF INVENTION:  Methods to Inhibit or Enhance the Binding of Viral DNA
;  TITLE OF INVENTION:  to Genomic Host DNA
;  FILE REFERENCE:  UM-03778
;  CURRENT APPLICATION NUMBER: US/09/410,399
;  CURRENT FILING DATE:  1999-10-01
;  NUMBER OF SEQ ID NOS:  6
;  SOFTWARE:  PatentIn Ver. 2.0
; SEQ ID NO 2
;   LENGTH: 1162
;   TYPE: PRT
;   ORGANISM: Kaposi's sarcoma-associated herpesvirus
US-09-410-399-2
```

Query Match 5.9%; Score 196.5; DB 2; Length 1162;
Best Local Similarity 24.2%; Pred. No. 1.2e-07;
Matches 96; Conservative 52; Mismatches 145; Indels 103; Gaps 14;

```
Qy        124 SSPYQDEEVISKTPTLAQLNSEDSQSVSDSLYYPDSLFSVKQNPLPSSFPGKKITSRAAA 183
              |||       :|   |        |: | | :|     |  | :: || ||  |:  :|
Db        111 SSPIPPSHPVSPGTT-------DTHSPSPALPPTQSPESSQRPPL-SSPTGRPDSSTPMR 162

Qy        184 PVCSSKTLQAEVPLSDCVQKASK------PPSSTQIMVKTNMYHNEKVNFHVECKDYVKK 237
              |  | :|     | :   :  ||       ||: : : |  :                 :
Db        163 PPPSQQTTPPHSPTTPPPEPPSKSSPDSLAPSTLRSLRKRRLSS--------------PQ 208

Qy        238 AKVKINPVQQSRPLL--------SQIHTDAAKENTCYCGAVAKRQEKKGMEPLQ----GH 285
                :||: || |:           ::   | |:  | |: :          |      |
Db        209 GPSTLNPICQSPPVSPPRCDFANRSVYPPWATESPIYVGSSSDGDTPPRQPPTSPISIGS 268

Qy        286 ATPALPF--KETQELLL--------SPLPQEGPGSLAAGESSSLSASTSVSDSSQKKEEH 335
              ::|:      :|  |:|        |   :|   :  |||    :   :|  ||  ::
Db        269 SSPSEGSWGDDTAMLVLLAEIAEEASKNEKECSENNQAGED---NGDNEISKESQVDKDD 325
```

```
Qy        336 NYSLFVSDNLGEQPTKCSPEEDEEDEE--------------------------DVDDED 368
              |       ||  ::  : : ||||||:|                          : |||:
Db        326 N------DNKDDEEEQETDEEDEEDDEEDDEEDDEEDDEEDDEEDDEEDDEEDDEEDDEE 379

Qy        369 HDEGFGSEHELSENEEEEEEEEDYEDDKDDDISDTFSEPGYENDSVEDLKE--------- 419
               ||    | :  |:|||:||||| ||| |:|  |       |:|  || ||
Db        380 DDEEDDEEEDEEEDEEEDEEEEDEEDDDDEDNEDE------EDDEEEDKKEDEEDGGDGN 433

Qy        420 -VTSISSRKRGKRRYFWEYSEQLTPSQQERMLRPSE 454
               || | :: :    |  :|  | ||| :  | :
Db        434 KTLSIQSSQQQQEPQQQE-PQQQEPQQQEPLQEPQQ 468


RESULT 8
US-09-894-273-2
; Sequence 2, Application US/09894273
; Patent No. 6756203
; GENERAL INFORMATION:
;  APPLICANT: Kieff, Elliott D.
;  APPLICANT:  Ballestas, Mary E.
;  APPLICANT:  Kaye, Kenneth M.
;  TITLE OF INVENTION: RHADINO VIRUS LANA ACTS IN TRANS ON A UNIT OF RHADINO
;  TITLE OF INVENTION:  VIRUS DNA TO MEDIATE EFFICIENT EPISOME PERSISTENCE
;  FILE REFERENCE: 16412-10001R
;  CURRENT APPLICATION NUMBER: US/09/894,273
;  CURRENT FILING DATE:  2001-06-28
;  PRIOR APPLICATION NUMBER: US 60/109,422
;  PRIOR FILING DATE: 1998-11-19
;  NUMBER OF SEQ ID NOS: 3
;  SOFTWARE: PatentIn Ver. 2.0
; SEQ ID NO 2
;   LENGTH: 1162
;   TYPE: PRT
;   ORGANISM: Kaposi's sarcoma-associated herpesvirus
US-09-894-273-2

  Query Match             5.9%;  Score 196.5;  DB 2;  Length 1162;
  Best Local Similarity   24.2%;  Pred. No. 1.2e-07;
  Matches   96;  Conservative   52;  Mismatches  145;  Indels  103;  Gaps   14;

Qy        124 SSPYQDEEVISKTPTLAQLNSEDSQSVSDSLYYPDSLFSVKQNPLPSSFPGKKITSRAAA 183
              |||      :|    |       |: | | :|    |  | :: || ||  |:  :|
Db        111 SSPIPPSHPVSPGTT-------DTHSPSPALPPTQSPESSQRPPL-SSPTGRPDSSTPMR 162

Qy        184 PVCSSKTLQAEVPLSDCVQKASK------PPSSTQIMVKTNMYHNEKVNFHVECKDYVKK 237
              |  | :|     | :   :  ||       ||: : : |  :                 :
Db        163 PPPSQQTTPPHSPTTPPPEPPSKSSPDSLAPSTLRSLRKRRLSS--------------PQ 208

Qy        238 AKVKINPVQQSRPLL--------SQIHTDAAKENTCYCGAVAKRQEKKGMEPLQ----GH 285
              :||: || |:          ::    | |:  | |: :       |        |
Db        209 GPSTLNPICQSPPVSPPRCDFANRSVYPPWATESPIYVGSSSDGDTPPRQPPTSPISIGS 268

Qy        286 ATPALPF--KETQELLL--------SPLPQEGPGSLAAGESSSLSASTSVSDSSQKKEEH 335
              ::|:     :|  |:|       |   :|   :  |||    :    :|  ||  ::
Db        269 SSPSEGSWGDDTAMLVLLAEIAEEASKNEKECSENNQAGED---NGDNEISKESQVDKDD 325

Qy        336 NYSLFVSDNLGEQPTKCSPEEDEEDEE--------------------------DVDDED 368
              |       ||  ::  : : ||||||:|                          : |||:
Db        326 N------DNKDDEEEQETDEEDEEDDEEDDEEDDEEDDEEDDEEDDEEDDEEDDEEDDEE 379

Qy        369 HDEGFGSEHELSENEEEEEEEEDYEDDKDDDISDTFSEPGYENDSVEDLKE--------- 419
               ||    | :  |:|||:||||| ||| |:|  |       |:|  || ||
Db        380 DDEEDDEEEDEEEDEEEDEEEDEEDDDDEDNEDE------EDDEEEDKKEDEEDGGDGN 433

Qy        420 -VTSISSRKRGKRRYFWEYSEQLTPSQQERMLRPSE 454
               || | :: :    |  :|  | ||| :  | :
Db        434 KTLSIQSSQQQQEPQQQE-PQQQEPQQQEPLQEPQQ 468


RESULT 9
US-08-375-300-4
; Sequence 4, Application US/08375300
; Patent No. 5679566
;  GENERAL INFORMATION:
;    APPLICANT:  Feng, He
;    APPLICANT:  Jacobson, Allan S.
;    TITLE OF INVENTION:  HETEROLOGOUS POLYPEPTIDE PRODUCTION IN
;    TITLE OF INVENTION:  THE ABSENCE OF NONSENSE-MEDIATED mRNA DECAY FUNCTION
;    NUMBER OF SEQUENCES:  6
;    CORRESPONDENCE ADDRESS:
;      ADDRESSEE:  Fish & Richardson
;      STREET:  225 Franklin Street Suite 3100
;      CITY:  Boston
;      STATE:  MA
;      COUNTRY:  USA
;      ZIP:  02110-2804
;    COMPUTER READABLE FORM:
;      MEDIUM TYPE:  Floppy disk
;      COMPUTER:  IBM PC compatible
;      OPERATING SYSTEM:  PC-DOS/MS-DOS
;      SOFTWARE:  PatentIn Release #1.0, Version #1.30
;    CURRENT APPLICATION DATA:
;      APPLICATION NUMBER:  US/08/375,300
;      FILING DATE:  20-JAN-1995
;    ATTORNEY/AGENT INFORMATION:
;      NAME:  Fasse, J. P.
;      REGISTRATION NUMBER:  32,983
;      REFERENCE/DOCKET NUMBER:  04020/046001
;    TELECOMMUNICATION INFORMATION:
;      TELEPHONE:  (617)542-5070
;      TELEFAX:  (617)542-8906
;      TELEX:  200154
;  INFORMATION FOR SEQ ID NO:  4:
;    SEQUENCE CHARACTERISTICS:
;      LENGTH:  764 amino acids
;      TYPE:  amino acid
;      STRANDEDNESS:  single
;      TOPOLOGY:  linear
;    MOLECULE TYPE:  protein
US-08-375-300-4

  Query Match             5.5%;  Score 184;  DB 1;  Length 764;
  Best Local Similarity   20.2%;  Pred. No. 7.4e-07;
  Matches   98;  Conservative   69;  Mismatches  159;  Indels  160;  Gaps   18;

Qy        133 ISKTPTLAQLNSEDSQSVSDSLYYPDSLFSVKQ--------NPLPSSFPGK-KITSRAAA 183
              |:  |   ||  |              ||: | ::          |  |::|  | |:  |
Db        382 INNQPNPFYLNYSDP---------PDNYFRIQLVTTILLNINRTPAAFTKKCKLLLR--- 429

Qy        184 PVCSSKTLQAEVPLSDCVQKASKPPSSTQIMVKT------NMYHNEKVNFHVECKDYVKK 237
                    |   | ||           |  |:  | :      |:: | |           ::
Db        430 -FFEYYTFIKEQPL----------PKETEFRVSSTFKKYENIFGNTK----------FER 468

Qy        238 AKVKINPVQQSRPLLSQIHTDAAKENTCYCGAVAKRQEKKGMEPLQGHATPALPFKETQE 297
              ::  :    :   ||  ::   :|::     : :    |:   |::          |  |
Db        469 SENLVESASRLESLLKSLNAIKSKDDRVKGSSASIHNGKESAVPIES-------ITEDDE 521

Qy        298 LLLSPLPQEGPGSLAAGESSSLSASTSVSDSSQKKEEHNYSLFVSDNLGEQPTKCSPEED 357
                  :   |    ||      ||  ::|:  | :                 :  :   |:|
Db        522 ---DEDDENDDGVDLLGEDEDAEISTPNTESAPGKHQ-----------AKQDESEDEDD 566

Qy        358 EEDEEDVDDEDHDEGFGSEHELSENEEEEEEEEDYEDDKDDDISDTFSEPGYENDSVEDL 417
              |:|:|| ||:| |:     |     | :|:::|::| ||| |::  |: |:  |   |
Db        567 EDDDEDDDDDDDDDDDGE----EGDEDDDEDDDDEDDDDEEEEDSDSDLEYGGDLDAD- 621

Qy        418 KEVTSISSRKRGKRRYFWEYSEQLTPSQQERMLRPSEWNRDTLPSNMYQKNGLHHGKYAV 477
                    |    :| : ||          || |:  |
Db        622 --------RDIEMKRMYEEY---------ERKLKDEE---------------------- 641

Qy        478 KKSRRTDVEDLTPNPKKLLQIGNELRKLNKVISDLTPVSELPLTAR----------PRSR 527
                  |   |:|    :|::|   : ||  ||::   ||   |:: :           | |
Db        642 ---ERKAEEELERQFQKMMQESIDARKSEKVVASKIPVISKPVSVQKPLLLKKSEEPSSS 698

Qy        528 KE-------KNKLAF-----------RACRL-----------KKKAQYEANKVKLWG 555
              ||         |:||          |  :|           ::| : | ||:|
Db        699 KETYEELSKPKKIAFTFLTKSGKKTQSRILQLPTDVKFVSDVLEEEEKLKTERNKIKKIV 758
```

```
Qy        556 LNTEYD 561
              |   :|
Db        759 LKRSFD 764


RESULT 10
US-09-177-431-4
; Sequence 4, Application US/09177431
; Patent No. 6071700
;  GENERAL INFORMATION:
;    APPLICANT:  He, Feng
;    APPLICANT:  Jacobson, Allan S.
;    TITLE OF INVENTION:  HETEROLOGOUS POLYPEPTIDE PRODUCTION IN THE
;    TITLE OF INVENTION:  ABSENCE OF NONSENSE-MEDIATED mRNA DECAY FUNCTION
;    NUMBER OF SEQUENCES:  18
;    CORRESPONDENCE ADDRESS:
;      ADDRESSEE:  Fish & Richardson P.C.
;      STREET:  225 Franklin Street
;      CITY:  Boston
;      STATE:  MA
;      COUNTRY:  USA
;      ZIP:  02110-2804
;    COMPUTER READABLE FORM:
;      MEDIUM TYPE:  Diskette
;      COMPUTER:  IBM Compatible
;      OPERATING SYSTEM:  Windows 95
;      SOFTWARE:  FastSEQ for Windows Version 2.0b
;    CURRENT APPLICATION DATA:
;      APPLICATION NUMBER:  US/09/177,431
;      FILING DATE:
;    PRIOR APPLICATION DATA:
;      APPLICATION NUMBER:  08/955,472
;      FILING DATE:
;    ATTORNEY/AGENT INFORMATION:
;      NAME:  Fasse, J. Peter
;      REGISTRATION NUMBER:  32,983
;      REFERENCE/DOCKET NUMBER:  07917/050001
;    TELECOMMUNICATION INFORMATION:
;      TELEPHONE:  617/542-5070
;      TELEFAX:  617/542-9806
;      TELEX:  200154
;  INFORMATION FOR SEQ ID NO:  4:
;    SEQUENCE CHARACTERISTICS:
;      LENGTH:  764 amino acids
;      TYPE:  amino acid
;      TOPOLOGY:  linear
;    MOLECULE TYPE:  peptide
US-09-177-431-4

  Query Match             5.5%;  Score 184;  DB 2;  Length 764;
  Best Local Similarity   20.2%;  Pred. No. 7.4e-07;
  Matches   98;  Conservative   69;  Mismatches  159;  Indels  160;  Gaps   18;

Qy        133 ISKTPTLAQLNSEDSQSVSDSLYYPDSLFSVKQ--------NPLPSSFPGK-KITSRAAA 183
              |:  |    ||  |              ||: | ::          |  |::|  | |:  |
Db        382 INNQPNPFYLNYSDP---------PDNYFRIQLVTTILLNINRTPAAFTKKCKLLLR--- 429

Qy        184 PVCSSKTLQAEVPLSDCVQKASKPPSSTQIMVKT------NMYHNEKVNFHVECKDYVKK 237
                  |   | ||           |  |:  | :      |:: | |           ::
Db        430 -FFEYYTFIKEQPL----------PKETEFRVSSTFKKYENIFGNTK----------FER 468

Qy        238 AKVKINPVQQSRPLLSQIHTDAAKENTCYCGAVAKRQEKKGMEPLQGHATPALPFKETQE 297
              ::  :    :   ||  ::   :|::     : :    |:   |::          | |
Db        469 SENLVESASRLESLLKSLNAIKSKDDRVKGSSASIHNGKESAVPIES-------ITEDDE 521

Qy        298 LLLSPLPQEGPGSLAAGESSSLSASTSVSDSSQKKEEHNYSLFVSDNLGEQPTKCSPEED 357
                  :   |    ||      ||  ::|:  | :          :  :    |:|
Db        522 ---DEDDENDDGVDLLGEDEDAEISTPNTESAPGKHQ-----------AKQDESEDEDD 566

Qy        358 EEDEEDVDDEDHDEGFGSEHELSENEEEEEEEEDYEDDKDDDISDTFSEPGYENDSVEDL 417
              |:|:|| ||:|  |:     |    | :|:::|::| ||| |::  |: |:  |  |   |
Db        567 EDDDEDDDDDDDDDDDGE----EGDEDDDEDDDDEDDDDEEEEDSDSDLEYGGDLDAD- 621

Qy        418 KEVTSISSRKRGKRRYFWEYSEQLTPSQQERMLRPSEWNRDTLPSNMYQKNGLHHGKYAV 477
                      |    :| : ||         || |:  |
Db        622 --------RDIEMKRMYEEY---------ERKLKDEE---------------------- 641

Qy        478 KKSRRTDVEDLTPNPKKLLQIGNELRKLNKVISDLTPVSELPLTAR----------PRSR 527
                  |   |:|    :|::|   : ||  ||::   ||   |:: :          | |
Db        642 ---ERKAEEELERQFQKMMQESIDARKSEKVVASKIPVISKPVSVQKPLLLKKSEEPSSS 698

Qy        528 KE-------KNKLAF-----------RACRL-------------KKKAQYEANKVKLWG 555
              ||         |:||           |  :|             ::| : | ||:|
Db        699 KETYEELSKPKKIAFTFLTKSGKKTQSRILQLPTDVKFVSDVLEEEEKLKTERNKIKKIV 758

Qy        556 LNTEYD 561
              |   :|
Db        759 LKRSFD 764


RESULT 11
PCT-US95-16930-4
; Sequence 4, Application PC/TUS9516930
;  GENERAL INFORMATION:
;    APPLICANT:  UNIVERSITY OF MASSACHUSETTS MEDICAL SCHOOL
;    TITLE OF INVENTION:  HETEROLOGOUS POLYPEPTIDE
;    TITLE OF INVENTION:  PRODUCTION IN THE ABSENCE OF
;    TITLE OF INVENTION:  NONSENSE-MEDIATED mRNA DECAY
;    TITLE OF INVENTION:  FUNCTION
;    NUMBER OF SEQUENCES:  6
;    CORRESPONDENCE ADDRESS:
;      ADDRESSEE:  Fish & Richardson
;      STREET:  225 Franklin Street Suite 3100
;      CITY:  Boston
;      STATE:  MA
;      COUNTRY:  USA
;      ZIP:  02110-2804
;    COMPUTER READABLE FORM:
;      MEDIUM TYPE:  Floppy disk
;      COMPUTER:  IBM PC compatible
;      OPERATING SYSTEM:  PC-DOS/MS-DOS
;      SOFTWARE:  PatentIn Release #1.0, Version #1.30
;    CURRENT APPLICATION DATA:
;      APPLICATION NUMBER:  PCT/US95/16930
;      FILING DATE:  27-DEC-1995
;    PRIOR APPLICATION DATA:
;      APPLICATION NUMBER:  US 08/375,300
;      FILING DATE:  20-JAN-1995
;    ATTORNEY/AGENT INFORMATION:
;      NAME:  Fasse, J. Peter
;      REGISTRATION NUMBER:  32,983
;      REFERENCE/DOCKET NUMBER:  04020/046WO1
;    TELECOMMUNICATION INFORMATION:
;      TELEPHONE:  (617)542-5070
;      TELEFAX:  (617)542-8906
;      TELEX:  200154
;  INFORMATION FOR SEQ ID NO:  4:
;    SEQUENCE CHARACTERISTICS:
;      LENGTH:  764 amino acids
;      TYPE:  amino acid
;      STRANDEDNESS:  single
;      TOPOLOGY:  linear
;    MOLECULE TYPE:  protein
PCT-US95-16930-4

  Query Match             5.5%;  Score 184;  DB 4;  Length 764;
  Best Local Similarity   20.2%;  Pred. No. 7.4e-07;
  Matches   98;  Conservative   69;  Mismatches  159;  Indels  160;  Gaps   18;

Qy        133 ISKTPTLAQLNSEDSQSVSDSLYYPDSLFSVKQ--------NPLPSSFPGK-KITSRAAA 183
              |:  |    ||  |              ||: | ::          |  |::|  | |:  |
Db        382 INNQPNPFYLNYSDP---------PDNYFRIQLVTTILLNINRTPAAFTKKCKLLLR--- 429

Qy        184 PVCSSKTLQAEVPLSDCVQKASKPPSSTQIMVKT------NMYHNEKVNFHVECKDYVKK 237
                  |   | ||           |  |:  | :      |:: | |           ::
```

```
Db        430 -FFEYYTFIKEQPL----------PKETEFRVSSTFKKYENIFGNTK----------FER 468

Qy        238 AKVKINPVQQSRPLLSQIHTDAAKENTCYCGAVAKRQEKKGMEPLQGHATPALPFKETQE 297

Db        469 SENLVESASRLESLLKSLNAIKSKDDRVKGSSASIHNGKESAVPIES-------ITEDDE 521

Qy        298 LLLSPLPQEGPGSLAAGESSSLSASTSVSDSSQKKEEHNYSLFVSDNLGEQPTKCSPEED 357

Db        522 ---DEDDENDDGVDLLGEDEDAEISTPNTESAPGKHQ-----------AKQDESEDEDD 566

Qy        358 EEDEEDVDDEDHDEGFGSEHELSENEEEEEEEEDYEDDKDDDISDTFSEPGYENDSVEDL 417

Db        567 EDDDEDDDDDDDDDDDDGE----EGDEDDDEDDDDEDDDDEEEEDSDSDLEYGGDLDAD- 621

Qy        418 KEVTSISSRKRGKRRYFWEYSEQLTPSQQERMLRPSEWNRDTLPSNMYQKNGLHHGKYAV 477

Db        622 --------RDIEMKRMYEEY---------ERKLKDEE---------------------- 641

Qy        478 KKSRRTDVEDLTPNPKKLLQIGNELRKLNKVISDLTPVSELPLTAR----------PRSR 527

Db        642 ---ERKAEEELERQFQKMMQESIDARKSEKVVASKIPVISKPVSVQKPLLLKKSEEPSSS 698

Qy        528 KE-------KNKLAF-----------RACRL-----------KKKAQYEANKVKLWG 555

Db        699 KETYEELSKPKKIAFTFLTKSGKKTQSRILQLPTDVKFVSDVLEEEEKLKTERNKIKKIV 758

Qy        556 LNTEYD 561

Db        759 LKRSFD 764
```

RESULT 12
US-08-375-300-2
; Sequence 2, Application US/08375300
; Patent No. 5679566
; GENERAL INFORMATION:
; APPLICANT: Feng, He
; APPLICANT: Jacobson, Allan S.
; TITLE OF INVENTION: HETEROLOGOUS POLYPEPTIDE PRODUCTION IN
; TITLE OF INVENTION: THE ABSENCE OF NONSENSE-MEDIATED mRNA DECAY FUNCTION
; NUMBER OF SEQUENCES: 6
; CORRESPONDENCE ADDRESS:
; ADDRESSEE: Fish & Richardson
; STREET: 225 Franklin Street Suite 3100
; CITY: Boston
; STATE: MA
; COUNTRY: USA
; ZIP: 02110-2804
; COMPUTER READABLE FORM:
; MEDIUM TYPE: Floppy disk
; COMPUTER: IBM PC compatible
; OPERATING SYSTEM: PC-DOS/MS-DOS
; SOFTWARE: PatentIn Release #1.0, Version #1.30
; CURRENT APPLICATION DATA:
; APPLICATION NUMBER: US/08/375,300
; FILING DATE: 20-JAN-1995
; ATTORNEY/AGENT INFORMATION:
; NAME: Fasse, J. P.
; REGISTRATION NUMBER: 32,983
; REFERENCE/DOCKET NUMBER: 04020/046001
; TELECOMMUNICATION INFORMATION:
; TELEPHONE: (617)542-5070
; TELEFAX: (617)542-8906
; TELEX: 200154
; INFORMATION FOR SEQ ID NO: 2:
; SEQUENCE CHARACTERISTICS:
; LENGTH: 1089 amino acids
; TYPE: amino acid
; STRANDEDNESS: not relevant
; TOPOLOGY: linear
; MOLECULE TYPE: protein
US-08-375-300-2
:

Query Match 5.5%; Score 184; DB 1; Length 1089;
Best Local Similarity 20.2%; Pred. No. 1.3e-06;
Matches 98; Conservative 69; Mismatches 159; Indels 160; Gaps 18;

```
Qy        133 ISKTPTLAQLNSEDSQSVSDSLYYPDSLFSVKQ--------NPLPSSFPGK-KITSRAAA 183

Db        707 INNQPNPFYLNYSDP---------PDNYFRIQLVTTILLNINRTPAAFTKKCKLLLR--- 754

Qy        184 PVCSSKTLQAEVPLSDCVQKASKPPSSTQIMVKT------NMYHNEKVNFHVECKDYVKK 237

Db        755 -FPEYYTFIKEQPL----------PKETEFRVSSTFKKYENIFGNTK----------FER 793

Qy        238 AKVKINPVQQSRPLLSQIHTDAAKENTCYCGAVAKRQEKKGMEPLQGHATPALPFKETQE 297

Db        794 SENLVESASRLESLLKSLNAIKSKDDRVKGSSASIHNGKESAVPIES-------ITEDDE 846

Qy        298 LLLSPLPQEGPGSLAAGESSSLSASTSVSDSSQKKEEHNYSLFVSDNLGEQPTKCSPEED 357

Db        847 ---DEDDENDDGVDLLGEDEDAEISTPNTESAPGKHQ-----------AKQDESEDEDD 891

Qy        358 EEDEEDVDDEDHDEGFGSEHELSENEEEEEEEEDYEDDKDDDISDTFSEPGYENDSVEDL 417

Db        892 EDDDEDDDDDDDDDDDDGE----EGDEDDDEDDDDEDDDDEEEEDSDSDLEYGGDLDAD- 946

Qy        418 KEVTSISSRKRGKRRYFWEYSEQLTPSQQERMLRPSEWNRDTLPSNMYQKNGLHHGKYAV 477

Db        947 --------RDIEMKRMYEEY---------ERKLKDEE---------------------- 966

Qy        478 KKSRRTDVEDLTPNPKKLLQIGNELRKLNKVISDLTPVSELPLTAR----------PRSR 527

Db        967 ---ERKAEEELERQFQKMMQESIDARKSEKVVASKIPVISKPVSVQKPLLLKKSEEPSSS 1023

Qy        528 KE-------KNKLAF-----------RACRL-----------KKKAQYEANKVKLWG 555

Db       1024 KETYEELSKPKKIAFTFLTKSGKKTQSRILQLPTDVKFVSDVLEEEEKLKTERNKIKKIV 1083

Qy        556 LNTEYD 561

Db       1084 LKRSFD 1089
```

RESULT 13
US-09-177-431-2
; Sequence 2, Application US/09177431
; Patent No. 6071700
; GENERAL INFORMATION:
; APPLICANT: He, Feng
; APPLICANT: Jacobson, Allan S.
; TITLE OF INVENTION: HETEROLOGOUS POLYPEPTIDE PRODUCTION IN THE
; TITLE OF INVENTION: ABSENCE OF NONSENSE-MEDIATED mRNA DECAY FUNCTION
; NUMBER OF SEQUENCES: 18
; CORRESPONDENCE ADDRESS:
; ADDRESSEE: Fish & Richardson P.C.
; STREET: 225 Franklin Street
; CITY: Boston
; STATE: MA
; COUNTRY: USA
; ZIP: 02110-2804
; COMPUTER READABLE FORM:
; MEDIUM TYPE: Diskette
; COMPUTER: IBM Compatible
; OPERATING SYSTEM: Windows 95
; SOFTWARE: FastSEQ for Windows Version 2.0b
; CURRENT APPLICATION DATA:
; APPLICATION NUMBER: US/09/177,431
; FILING DATE:
; PRIOR APPLICATION DATA:
; APPLICATION NUMBER: 08/955,472
; FILING DATE:
; ATTORNEY/AGENT INFORMATION:
; NAME: Fasse, J. Peter
; REGISTRATION NUMBER: 32,983
; REFERENCE/DOCKET NUMBER: 07917/050001

```
;    TELECOMMUNICATION INFORMATION:
;      TELEPHONE:  617/542-5070
;      TELEFAX:  617/542-9806
;      TELEX:  200154
;  INFORMATION FOR SEQ ID NO:  2:
;    SEQUENCE CHARACTERISTICS:
;      LENGTH:  1089 amino acids
;      TYPE:  amino acid
;      TOPOLOGY:  linear
;    MOLECULE TYPE:  protein
US-09-177-431-2

  Query Match             5.5%;  Score 184;  DB 2;  Length 1089;
  Best Local Similarity   20.2%;  Pred. No. 1.3e-06;
  Matches   98;  Conservative   69;  Mismatches  159;  Indels  160;  Gaps   18;

Qy        133 ISKTPTLAQLNSEDSQSVSDSLYYPDSLFSVKQ--------NPLPSSFPGK-KITSRAAA 183
              |:  |    ||  |          ||: | ::          |  |::|  | |:  |
Db        707 INNQPNPFYLNYSDP---------PDNYFRIQLVTTILLNINRTPAAFTKKCKLLLR--- 754

Qy        184 PVCSSKTLQAEVPLSDCVQKASKPPSSTQIMVKT------NMYHNEKVNFHVECKDYVKK 237
                  |   | ||           |  |:  | :      |:: | |           ::
Db        755 -FFEYYTFIKEQPL----------PKETEFRVSSTFKKYENIFGNTK----------FER 793

Qy        238 AKVKINPVQQSRPLLSQIHTDAAKENTCYCGAVAKRQEKKGMEPLQGHATPALPFKETQE 297
              ::  :    :   ||  ::   :|::     : :    |:   |::           | |
Db        794 SENLVESASRLESLLKSLNAIKSKDDRVKGSSASIHNGKESAVPIES-------ITEDDE 846

Qy        298 LLLSPLPQEGPGSLAAGESSSLSASTSVSDSSQKKEEHNYSLFVSDNLGEQPTKCSPEED 357
                  :   |    ||      ||  ::|:  | :          :  :    |:|
Db        847 ---DEDDENDDGVDLLGEDEDAEISTPNTESAPGKHQ-----------AKQDESEDEDD 891

Qy        358 EEDEEDVDDEDHDEGFGSEHELSENEEEEEEEEEDYEDDKDDDISDTFSEPGYENDSVEDL 417
              |:|:|| ||:| |:     |    | :|:::|::| ||| |::  |: |:  | |   |
Db        892 EDDDEDDDDDDDDDDDDGE----EGDEDDDEDDDDEDDDDEEEEDSDSDLEYGGDLDAD- 946

Qy        418 KEVTSISSRKRGKRRYFWEYSEQLTPSQQERMLRPSEWNRDTLPSNMYQKNGLHHGKYAV 477
                    |    :| : ||          || |:  |
Db        947 --------RDIEMKRMYEEY---------ERKLKDEE---------------------- 966

Qy        478 KKSRRTDVEDLTPNPKKLLQIGNELRKLNKVISDLTPVSELPLTAR----------PRSR 527
                |   |:|    :|::|   : ||  ||::   ||   |:: :          | |
Db        967 ---ERKAEEELERQFQKMMQESIDARKSEKVVASKIPVISKPVSVQKPLLLKKSEEPSSS 1023

Qy        528 KE-------KNKLAF-----------RACRL-------------KKKAQYEANKVKLWG 555
              ||          |:||          |  :|             ::| : | ||:|
Db       1024 KETYEELSKPKKIAFTFLTKSGKKTQSRILQLPTDVKFVSDVLEEEEKLKTERNKIKKIV 1083

Qy        556 LNTEYD 561
              |  :|
Db       1084 LKRSFD 1089


RESULT 14
PCT-US95-16930-2
; Sequence 2, Application PC/TUS9516930
;  GENERAL INFORMATION:
;    APPLICANT:  UNIVERSITY OF MASSACHUSETTS MEDICAL SCHOOL
;    TITLE OF INVENTION:  HETEROLOGOUS POLYPEPTIDE
;    TITLE OF INVENTION:  PRODUCTION IN THE ABSENCE OF
;    TITLE OF INVENTION:  NONSENSE-MEDIATED mRNA DECAY
;    TITLE OF INVENTION:  FUNCTION
;    NUMBER OF SEQUENCES:  6
;    CORRESPONDENCE ADDRESS:
;      ADDRESSEE:  Fish & Richardson
;      STREET:  225 Franklin Street Suite 3100
;      CITY:  Boston
;      STATE:  MA
;      COUNTRY:  USA
;      ZIP:  02110-2804
;    COMPUTER READABLE FORM:
;      MEDIUM TYPE:  Floppy disk
;      COMPUTER:  IBM PC compatible
;      OPERATING SYSTEM:  PC-DOS/MS-DOS
;      SOFTWARE:  PatentIn Release #1.0, Version #1.30
;    CURRENT APPLICATION DATA:
;      APPLICATION NUMBER:  PCT/US95/16930
;      FILING DATE:  27-DEC-1995
;    PRIOR APPLICATION DATA:
;      APPLICATION NUMBER:  US 08/375,300
;      FILING DATE:  20-JAN-1995
;    ATTORNEY/AGENT INFORMATION:
;      NAME:  Fasse, J. Peter
;      REGISTRATION NUMBER:  32,983
;      REFERENCE/DOCKET NUMBER:  04020/046WO1
;    TELECOMMUNICATION INFORMATION:
;      TELEPHONE:  (617)542-5070
;      TELEFAX:  (617)542-8906
;      TELEX:  200154
;  INFORMATION FOR SEQ ID NO:  2:
;    SEQUENCE CHARACTERISTICS:
;      LENGTH:  1089 amino acids
;      TYPE:  amino acid
;      STRANDEDNESS:  not relevant
;      TOPOLOGY:  linear
;    MOLECULE TYPE:  protein
PCT-US95-16930-2

  Query Match             5.5%;  Score 184;  DB 4;  Length 1089;
  Best Local Similarity   20.2%;  Pred. No. 1.3e-06;
  Matches   98;  Conservative   69;  Mismatches  159;  Indels  160;  Gaps   18;

Qy        133 ISKTPTLAQLNSEDSQSVSDSLYYPDSLFSVKQ--------NPLPSSFPGK-KITSRAAA 183
              |:  |    ||  |          ||: | ::          |  |::|  | |:  |
Db        707 INNQPNPFYLNYSDP---------PDNYFRIQLVTTILLNINRTPAAFTKKCKLLLR--- 754

Qy        184 PVCSSKTLQAEVPLSDCVQKASKPPSSTQIMVKT------NMYHNEKVNFHVECKDYVKK 237
                  |   | ||           |  |:  | :      |:: | |           ::
Db        755 -FFEYYTFIKEQPL----------PKETEFRVSSTFKKYENIFGNTK----------FER 793

Qy        238 AKVKINPVQQSRPLLSQIHTDAAKENTCYCGAVAKRQEKKGMEPLQGHATPALPFKETQE 297
              ::  :    :   ||  ::   :|::     : :    |:   |::           | |
Db        794 SENLVESASRLESLLKSLNAIKSKDDRVKGSSASIHNGKESAVPIES-------ITEDDE 846

Qy        298 LLLSPLPQEGPGSLAAGESSSLSASTSVSDSSQKKEEHNYSLFVSDNLGEQPTKCSPEED 357
                  :   |    ||      ||  ::|:  | :          :  :    |:|
Db        847 ---DEDDENDDGVDLLGEDEDAEISTPNTESAPGKHQ-----------AKQDESEDEDD 891

Qy        358 EEDEEDVDDEDHDEGFGSEHELSENEEEEEEEEEDYEDDKDDDISDTFSEPGYENDSVEDL 417
              |:|:|| ||:| |:     |    | :|:::|::| ||| |::  |: |:  | |   |
Db        892 EDDDEDDDDDDDDDDDDGE----EGDEDDDEDDDDEDDDDEEEEDSDSDLEYGGDLDAD- 946

Qy        418 KEVTSISSRKRGKRRYFWEYSEQLTPSQQERMLRPSEWNRDTLPSNMYQKNGLHHGKYAV 477
                    |    :| : ||          || |:  |
Db        947 --------RDIEMKRMYEEY---------ERKLKDEE---------------------- 966

Qy        478 KKSRRTDVEDLTPNPKKLLQIGNELRKLNKVISDLTPVSELPLTAR----------PRSR 527
                |   |:|    :|::|   : ||  ||::   ||   |:: :          | |
Db        967 ---ERKAEEELERQFQKMMQESIDARKSEKVVASKIPVISKPVSVQKPLLLKKSEEPSSS 1023

Qy        528 KE-------KNKLAF-----------RACRL-------------KKKAQYEANKVKLWG 555
              ||          |:||          |  :|             ::| : | ||:|
Db       1024 KETYEELSKPKKIAFTFLTKSGKKTQSRILQLPTDVKFVSDVLEEEEKLKTERNKIKKIV 1083

Qy        556 LNTEYD 561
              |  :|
Db       1084 LKRSFD 1089


RESULT 15
US-09-248-796A-15197
; Sequence 15197, Application US/09248796A
; Patent No. 6747137
; GENERAL INFORMATION:
;  APPLICANT: Keith Weinstock et al
;  TITLE OF INVENTION: NUCLEIC ACID AND AMINO ACID SEQUENCES RELATING TO CANDIDA ALBICANS
```

```
;  TITLE OF INVENTION:  FOR DIAGNOSTICS AND THERAPEUTICS
;  FILE REFERENCE: 107196.132
;  CURRENT APPLICATION NUMBER: US/09/248,796A
;  CURRENT FILING DATE:  1999-02-12
;  PRIOR APPLICATION NUMBER: US 60/074,725
;  PRIOR FILING DATE: 1998-02-13
;  PRIOR APPLICATION NUMBER: US 60/096,409
;  PRIOR FILING DATE: 1998-08-13
;  NUMBER OF SEQ ID NOS: 28208
; SEQ ID NO 15197
;   LENGTH: 428
;   TYPE: PRT
;   ORGANISM: Candida albicans
US-09-248-796A-15197

  Query Match             5.3%;  Score 176.5;  DB 2;  Length 428;
  Best Local Similarity   21.7%;  Pred. No. 1.4e-06;
  Matches  103;  Conservative   71;  Mismatches  152;  Indels  149;  Gaps   19;

Qy        197 LSDCVQKA-SKPPSSTQIMVKTNMYHNEKVNFHVECKDYVKKAKV--KINPVQQSRPLLS 253
              | | : || |||   | :  ||           | |: :: | |  |:   ::|:
Db         17 LIDAIDKAKSKPTKKTPVTEKT----------VEEEKEIIEAAPVTTKVKSNKKSK---- 62

Qy        254 QIHTDAAKENTCYCGAVAKRQEKKGMEPLQGHATPALPFKETQELLLSPLPQEGPGSLAA 313
                 ::  |::  |   |  ::|:: :: ||         |: :|
Db         63 --KSNTKKDDDDYQSEVLSKKEQRRLKKLQ--------TKQQEE---------------- 96

Qy        314 GESSSLSASTSVSDSSQKKEEHNYSLFVSDNLGEQPTKCSPEEDEEDEEDVDDEDHDEGF 373
                          :|: |:|                    ||:|:|::| :: | ::
Db         97 ---------------EKEAENN---------------EEEEEEDDDDDEEELDLEKLA 124

Qy        374 GSEHELSENEEEEE------EEEDYEDDKDDDISDTFSEPGYENDSVEDLKEVTSISSRK 427
              || |   |:||||      |||| | : ||:  |       | ||  |:   | ::
Db        125 ASESESDINDEEEEDDDDDEEEDDEVENDDEAEDDIPLSDVEVDSDADIVPHTKLTINN 184

Qy        428 RG-------------KRRYFWEYSEQLTPSQQERMLRPSEWNRDTLPSNMYQKNGLHHGK 474
                           :  | |:    :  : |  ::  :   ||     : | ||
Db        185 MAALRESLARIELPWSKHSFIEHQSITSADKTESEIK--DIYDDTERELAFYKQGLD--- 239

Qy        475 YAVKKSRRTDVEDLTPNPKKLLQIGNELRKLNKVISDLTPVSELPLTARPRSRKEKNKLA 534
              |||:||:|          ||::     :     :::    |          | ||||
Db        240 -AVKQSRKT----------LLKLKIPFSRPMDYFAEMVKSDE-------HMDKLKNKLL 280

Qy        535 FRACRLKKKAQYEANK---VKLWGLNTEYDNLLFVINSIKQEIVNRVQNPRDERGPNMGQ 591
              |    |||  || :   :| :|   ::  |       |:| : :::: : :|| |
Db        281 TEAA--NKKASEEAKRQRQLKKFGKQVQHATLQERAKQ-KKETLEKIKSLKKKRGAN--- 334

Qy        592 KLEILIKDTLGLPVAGQTSEFVNQVLEKTAE--------GNPTGGLVGLRIPTSK 638
              ||   |   :              ||:  |        |: :||    | | ||
Db        335 --EISNDDDFQI------------ALEEATENNQKYGHGGSGSGGDNKRRKPNSK 375


Search completed: February 28, 2006, 08:56:05
Job time : 56.235 secs
```

OM protein - protein search, using sw model

Run on: February 28, 2006, 08:54:36 ; Search time 186.308 Seconds
(without alignments)
1433.071 Million cell updates/sec

Title: US-10-717-665A-44
Perfect score: 3349
Sequence: 1 MPQPSVSGMDPPFGDAFRSH.....TAEGNPTGGLVGLRIPTSKV 639

Scoring table: BLOSUM62
Gapop 10.0 , Gapext 0.5

Searched: 1867569 seqs, 417829326 residues

Total number of hits satisfying chosen parameters: 1867569

Minimum DB seq length: 0
Maximum DB seq length: 2000000000

Post-processing: Minimum Match 0%
Maximum Match 100%
Listing first 45 summaries

Database : Published_Applications_AA_Main:*
1: /cgn2_6/ptodata/1/pubpaa/US07_PUBCOMB.pep:*
2: /cgn2_6/ptodata/1/pubpaa/US08_PUBCOMB.pep:*
3: /cgn2_6/ptodata/1/pubpaa/US09_PUBCOMB.pep:*
4: /cgn2_6/ptodata/1/pubpaa/US10A_PUBCOMB.pep:*
5: /cgn2_6/ptodata/1/pubpaa/US10B_PUBCOMB.pep:*
6: /cgn2_6/ptodata/1/pubpaa/US11_PUBCOMB.pep:*

Pred. No. is the number of results predicted by chance to have a score greater than or equal to the score of the result being printed, and is derived by analysis of the total score distribution.

SUMMARIES

Result No.	Score	% Query Match	Length	DB	ID	Description
1	3349	100.0	639	5	US-10-717-665-44	Sequence 44, Appl
2	506	15.1	814	6	US-11-097-143-31914	Sequence 31914, A
3	434	13.0	91	5	US-10-717-665-75	Sequence 75, Appl
4	200.5	6.0	160	5	US-10-450-763-59634	Sequence 59634, A
5	196.5	5.9	1162	3	US-09-894-273-2	Sequence 2, Appli
6	196.5	5.9	1162	4	US-10-294-804-2	Sequence 2, Appli
7	196.5	5.9	1162	5	US-10-194-046-2	Sequence 2, Appli
8	194.5	5.8	1003	5	US-10-732-923-3319	Sequence 3319, Ap
9	189.5	5.7	735	5	US-10-483-505-3	Sequence 3, Appli
10	179	5.3	427	4	US-10-032-585-7386	Sequence 7386, Ap
11	175	5.2	2724	5	US-10-487-593-1	Sequence 1, Appli
12	174	5.2	707	3	US-09-825-886-22	Sequence 22, Appl
13	174	5.2	707	4	US-10-384-569-3	Sequence 3, Appli
14	174	5.2	707	4	US-10-400-083-19	Sequence 19, Appl
15	174	5.2	707	5	US-10-491-545A-20	Sequence 20, Appl
16	173.5	5.2	2097	5	US-10-450-763-40303	Sequence 40303, A
17	170	5.1	513	4	US-10-374-780A-522	Sequence 522, App
18	170	5.1	513	4	US-10-389-566-1147	Sequence 1147, Ap
19	170	5.1	1915	5	US-10-661-398-36	Sequence 36, Appl
20	166.5	5.0	652	4	US-10-425-114-58474	Sequence 58474, A
21	166.5	5.0	960	4	US-10-128-714-8193	Sequence 8193, Ap
22	166.5	5.0	997	4	US-10-320-797-3328	Sequence 3328, Ap
23	166	5.0	1579	5	US-10-504-582-158	Sequence 158, App
24	165.5	4.9	1502	4	US-10-369-493-3947	Sequence 3947, Ap
25	165	4.9	777	5	US-10-450-763-49448	Sequence 49448, A
26	165	4.9	920	6	US-11-097-143-29151	Sequence 29151, A
27	164.5	4.9	687	4	US-10-104-047-2651	Sequence 2651, Ap
28	164.5	4.9	687	6	US-11-013-684-7	Sequence 7, Appli
29	163.5	4.9	802	4	US-10-369-493-22462	Sequence 22462, A
30	163.5	4.9	802	4	US-10-451-467A-388	Sequence 388, App
31	163.5	4.9	1972	3	US-09-839-479-21	Sequence 21, Appl
32	163.5	4.9	1972	4	US-10-376-537-21	Sequence 21, Appl
33	163.5	4.9	1972	4	US-10-702-148-21	Sequence 21, Appl
34	163.5	4.9	1972	5	US-10-489-740-176	Sequence 176, App
35	163.5	4.9	1972	5	US-10-756-149-5621	Sequence 5621, Ap
36	162.5	4.9	714	3	US-09-978-242-3	Sequence 3, Appli
37	162.5	4.9	1441	5	US-10-732-923-3352	Sequence 3352, Ap
38	162	4.8	167	5	US-10-450-763-41629	Sequence 41629, A
39	162	4.8	3328	5	US-10-732-923-8311	Sequence 8311, Ap
40	161.5	4.8	542	4	US-10-218-137-21	Sequence 21, Appl
41	161	4.8	1468	6	US-11-097-143-15765	Sequence 15765, A
42	160.5	4.8	382	5	US-10-450-763-34710	Sequence 34710, A
43	160.5	4.8	1969	3	US-09-839-479-71	Sequence 71, Appl
44	160.5	4.8	1969	4	US-10-376-537-72	Sequence 72, Appl
45	160.5	4.8	1969	4	US-10-702-148-71	Sequence 71, Appl

ALIGNMENTS

RESULT 1
US-10-717-665-44
; Sequence 44, Application US/10717665
; Publication No. US20050106579A1
; GENERAL INFORMATION:
; APPLICANT: OriGene Technologies, Inc
; TITLE OF INVENTION: Regulated Angiogenesis Genes and Polypeptides
; FILE REFERENCE: 1U 103 R1
; CURRENT APPLICATION NUMBER: US/10/717,665
; CURRENT FILING DATE: 2003-11-21
; PRIOR APPLICATION NUMBER: US/10/164,595
; PRIOR FILING DATE: 2002-06-10
; NUMBER OF SEQ ID NOS: 80
; SOFTWARE: PatentIn version 3.1
; SEQ ID NO 44
; LENGTH: 639
; TYPE: PRT
; ORGANISM: Homo sapiens
US-10-717-665-44

Query Match 100.0%; Score 3349; DB 5; Length 639;
Best Local Similarity 100.0%; Pred. No. 3.1e-211;
Matches 639; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

```
Qy          1 MPQPSVSGMDPPFGDAFRSHTFSEQTLMSTDLLANSSDPDFMYELDREMNYQQNPRDNFL 60
              ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db          1 MPQPSVSGMDPPFGDAFRSHTFSEQTLMSTDLLANSSDPDFMYELDREMNYQQNPRDNFL 60

Qy         61 SLEDCKDIENLESFTDVLDNEGALTSNWEQWDTYCEDLTKYTKLTSCDIWGTKEVDYLGL 120
              ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db         61 SLEDCKDIENLESFTDVLDNEGALTSNWEQWDTYCEDLTKYTKLTSCDIWGTKEVDYLGL 120

Qy        121 DDFSSPYQDEEVISKTPTLAQLNSEDSQSVSDSLYYPDSLFSVKQNPLPSSFPGKKITSR 180
              ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db        121 DDFSSPYQDEEVISKTPTLAQLNSEDSQSVSDSLYYPDSLFSVKQNPLPSSFPGKKITSR 180

Qy        181 AAAPVCSSKTLQAEVPLSDCVQKASKPPSSTQIMVKTNMYHNEKVNFHVECKDYVKKAKV 240
              ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db        181 AAAPVCSSKTLQAEVPLSDCVQKASKPPSSTQIMVKTNMYHNEKVNFHVECKDYVKKAKV 240

Qy        241 KINPVQQSRPLLSQIHTDAAKENTCYCGAVAKRQEKKGMEPLQGHATPALPFKETQELLL 300
              ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db        241 KINPVQQSRPLLSQIHTDAAKENTCYCGAVAKRQEKKGMEPLQGHATPALPFKETQELLL 300

Qy        301 SPLPQEGPGSLAAGESSSLSASTSVSDSSQKKEEHNYSLFVSDNLGEQPTKCSPEEDEED 360
              ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db        301 SPLPQEGPGSLAAGESSSLSASTSVSDSSQKKEEHNYSLFVSDNLGEQPTKCSPEEDEED 360

Qy        361 EEDVDDEDHDEGFGSEHELSENEEEEEEEEDYEDDKDDDISDTFSEPGYENDSVEDLKEV 420
              ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
```

```
Db        361 EEDVDDEDHDEGFGSEHELSENEEEEEEEEDYEDDKDDDISDTFSEPGYENDSVEDLKEV 420

Qy        421 TSISSRKRGKRRYFWEYSEQLTPSQQERMLRPSEWNRDTLPSNMYQKNGLHHGKYAVKKS 480
              ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db        421 TSISSRKRGKRRYFWEYSEQLTPSQQERMLRPSEWNRDTLPSNMYQKNGLHHGKYAVKKS 480

Qy        481 RRTDVEDLTPNPKKLLQIGNELRKLNKVISDLTPVSELPLTARPRSRKEKNKLAFRACRL 540
              ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db        481 RRTDVEDLTPNPKKLLQIGNELRKLNKVISDLTPVSELPLTARPRSRKEKNKLAFRACRL 540

Qy        541 KKKAQYEANKVKLWGLNTEYDNLLFVINSIKQEIVNRVQNPRDERGPNMGQKLEILIKDT 600
              ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db        541 KKKAQYEANKVKLWGLNTEYDNLLFVINSIKQEIVNRVQNPRDERGPNMGQKLEILIKDT 600

Qy        601 LGLPVAGQTSEFVNQVLEKTAEGNPTGGLVGLRIPTSKV 639
              ||||||||||||||||||||||||||||||||||||||
Db        601 LGLPVAGQTSEFVNQVLEKTAEGNPTGGLVGLRIPTSKV 639


RESULT 2
US-11-097-143-31914
; Sequence 31914, Application US/11097143
; Publication No. US20050208558A1
; GENERAL INFORMATION:
;  APPLICANT: Venter, J. Craig
;  APPLICANT:  et al.
;  TITLE OF INVENTION: DETECTION KIT, SUCH AS NUCLEIC ACID
;  TITLE OF INVENTION:  ARRAYS, FOR DETECTING EXPRESSION OF 10,000 OR MORE
;  TITLE OF INVENTION:  DROSOPHILA GENES.
;  FILE REFERENCE: CL000728
;  CURRENT APPLICATION NUMBER: US/11/097,143
;  CURRENT FILING DATE:  2005-04-04
;  PRIOR APPLICATION NUMBER: 60/157,832
;  PRIOR FILING DATE: 1999-10-05
;  PRIOR APPLICATION NUMBER: 60/160,191
;  PRIOR FILING DATE: 1999-10-19
;  PRIOR APPLICATION NUMBER: 60/161,932
;  PRIOR FILING DATE: 1999-10-28
;  PRIOR APPLICATION NUMBER: 60/164,769
;  PRIOR FILING DATE: 1999-11-12
;  PRIOR APPLICATION NUMBER: 60/173,383
;  PRIOR FILING DATE: 1999-12-28
;  PRIOR APPLICATION NUMBER: 60/175,693
;  PRIOR FILING DATE: 2000-01-12
;  PRIOR APPLICATION NUMBER: 60/184,831
;  PRIOR FILING DATE: 2000-02-24
;  PRIOR APPLICATION NUMBER: 60/191,637
;  PRIOR FILING DATE: 2000-03-23
;  NUMBER OF SEQ ID NOS: 43008
;  SOFTWARE: FastSEQ for Windows Version 4.0
; SEQ ID NO 31914
;   LENGTH: 814
;   TYPE: PRT
;   ORGANISM: DROSOPHILA
US-11-097-143-31914

  Query Match             15.1%;  Score 506;  DB 6;  Length 814;
  Best Local Similarity   25.0%;  Pred. No. 1.7e-24;
  Matches  186;  Conservative   94;  Mismatches  211;  Indels  254;  Gaps   25;

Qy          5 SVSGMDPPFGDAF---RSHTFSEQTLMSTDLLANSSDPDFMYELDREMNYQQNPRDNFL- 60
              | :|:    |  |    ||: |   :    ||  :: |:       |: |:  |  | :
Db        205 SSNGIGSGSGGYFTPDMSHSLSLNVVSEQVLLQEATTPN-------ELLYEMTPNSNAMW 257

Qy         61 ---------------SLED------------CKDIENLE--SFTDVLDNEGALTSNWEQW 91
                           ||:|               |: :|    | ||: |
Db        258 SDISSAIIHTKHEPFSLDDDYIFPNDKAEIQAADLSDLNGGDFLDVIGN----------- 306

Qy         92 DTYCEDLTKYTKLTSCDIWGTKEVDYL------GLDDFSSP-----------YQDEEV-- 132
                 ||    | :       |: |::|      | |   :|           :|  :|
Db        307 ---IEDFLPQTAV-------TQSVNFLLSPQAQGQDALVAPPMELLQQQQQNHQQLQVGS 356

Qy        133 ---ISKTPTLAQLNSEDSQSVSDSLYYPDSLFSVKQNPLPSSF-PGKKITSRAAAPVCSS 188
                  :    ||:|    :| | |    |  :    :|  | :|| || : :      |
Db        357 LPQLQTLLTLSQQQQSNSSSTSPYEIYHSTPQKPQQQQLSASFSPGSQASQSPLTP---- 412

Qy        189 KTLQAEVPLSDCVQKASKPPSST----QIMVKTNMYHNEKVNFHVECKDYVKKAKVKINP 244
                                 || :     |::   ||            :: :||     :|
Db        413 ---------------PPPPHANRPQYQMVKSRNM-----------QELIKKGFPMSSP 444

Qy        245 VQQSRPLLSQIHTDAAKENTCYCGAVAKRQEKKGMEPLQGHATPALPFKETQELLLSPLP 304
              ::|  :||                                                |  |||
Db        445 PERS--ILS-----------------------------------------QSAALSPGG 460

Qy        305 QEGPGSLAAGESSSLSASTSVSDSSQKKEEHNYSLFVSDNLGEQPTKCSPEEDEEDEEDV 364
                | || |:| |:: |  ||    |  ::   |   | ::   : :  :|
Db        461 SSGFGSSASGNSTTTSNQTS---GSAVRKSFGYQSAVENSQLSRLSSSAPTH-------- 509

Qy        365 DDEDHDEGFGSEH---------------ELSENEE----------------EEEEEEDY 392
                  | ||                   |:| |                    :::|:|
Db        510 --------LGLEHIWMRREPRQHLLSTGSLAEAESFSSLSTGSVLSPDGIDFSQDDEDDN 561

Qy        393 EDDKDDDISDTFSEPGYENDSVEDLKEVTSISSRKRGKRRYFWEYSEQLTPSQQERML-- 450
              :  |:  |  |: |   |  |      :  |  :|| |:||:|: |    : :|::
Db        562 SSENSDNYDDCSSDNGLSEDEDETRTSTPNHLSSSKGKERFFWQYNVQAKGPKGKRLVFQ 621

Qy        451 ----RPSEWNRDTLP----------SNMYQKNGLHHGKYAVKKSRRTDVEDLTPNPKKLL 496
                   |   | | |           :|: :|         |:|: |  ||||| :||
Db        622 SKLEDPHVLNEVTDPVFSPTCSVRGIKVYKHSG---------KARKGDGNDLTPNARKLH 672

Qy        497 QIGNELRKLNKVISDLTPVSELPLTARPRSRKEKNKLAFRACRLKKKAQYEANKVKLWGL 556
              || || ||:: |:|:|||||||   ||:|||||||| ||||||||||:||||:||:||
Db        673 NIGKELDKLSRTINDMTPVSELPFNVRPKSRKEKNKLASRACRLKKKAQHEANKIKLFGL 732

Qy        557 NTEYDNLLFVINSIKQEIVNRVQNPRDERGPNMGQKLE--------ILIKDTLGLPVAGQ 608
              |:  |:  |  :|| :| :       |  |:|:  |          |    : |: :||
Db        733 EIEHKRLMNGIAELKQALVVK------HRTKNLGESTEEVDQQIARIYATASSGIRIAGG 786

Qy        609 TSEFVNQVLEKTAEGNPTGGLVGLR 633
              :::|||:|||    | | |||  ||
Db        787 STDFVNKVLENMRGGMPNGGLEELR 811


RESULT 3
US-10-717-665-75
; Sequence 75, Application US/10717665
; Publication No. US20050106579A1
; GENERAL INFORMATION:
;  APPLICANT: OriGene Technologies, Inc
;  TITLE OF INVENTION: Regulated Angiogenesis Genes and Polypeptides
;  FILE REFERENCE: 1U 103 R1
;  CURRENT APPLICATION NUMBER: US/10/717,665
;  CURRENT FILING DATE:  2003-11-21
;  PRIOR APPLICATION NUMBER: US/10/164,595
;  PRIOR FILING DATE: 2002-06-10
;  NUMBER OF SEQ ID NOS: 80
;  SOFTWARE: PatentIn version 3.1
; SEQ ID NO 75
;   LENGTH: 91
;   TYPE: PRT
;   ORGANISM: Homo sapiens
US-10-717-665-75

  Query Match             13.0%;  Score 434;  DB 5;  Length 91;
  Best Local Similarity   100.0%;  Pred. No. 5.2e-21;
  Matches   83;  Conservative    0;  Mismatches    0;  Indels    0;  Gaps    0;

Qy        449 MLRPSEWNRDTLPSNMYQKNGLHHGKYAVKKSRRTDVEDLTPNPKKLLQIGNELRKLNKV 508
              |||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db          1 MLRPSEWNRDTLPSNMYQKNGLHHGKYAVKKSRRTDVEDLTPNPKKLLQIGNELRKLNKV 60

Qy        509 ISDLTPVSELPLTARPRSRKEKN 531
              |||||||||||||||||||||||
Db         61 ISDLTPVSELPLTARPRSRKEKN 83
```

```
RESULT 4
US-10-450-763-59634
; Sequence 59634, Application US/10450763
; Publication No. US20050196754A1
; GENERAL INFORMATION:
;  APPLICANT: Hyseq, Inc
;  TITLE OF INVENTION: NOVEL NUCLEIC ACIDS AND POLYPEPTIDES
;  FILE REFERENCE: 790CIP3/US
;  CURRENT APPLICATION NUMBER: US/10/450,763
;  CURRENT FILING DATE:  2003-06-11
;  PRIOR APPLICATION NUMBER: PCT/US01/08631
;  PRIOR FILING DATE: 2001-03-30
;  PRIOR APPLICATION NUMBER: 09/540,217
;  PRIOR FILING DATE: 2000-03-31
;  PRIOR APPLICATION NUMBER: 09/649,167
;  PRIOR FILING DATE: 2000-08-23
;  NUMBER OF SEQ ID NOS: 60736
;  SOFTWARE: Custom
; SEQ ID NO 59634
;   LENGTH: 160
;   TYPE: PRT
;   ORGANISM: Homo sapiens
US-10-450-763-59634

  Query Match             6.0%;  Score 200.5;  DB 5;  Length 160;
  Best Local Similarity   73.0%;  Pred. No. 2.3e-05;
  Matches   46;  Conservative    2;  Mismatches   12;  Indels    3;  Gaps    2;

Qy        540 LKKKAQYEANKVKLW-GLNTEYDNLLFVINSIKQEIVNRVQNPRDERGPNMGQKLEILIK 598
              |  |   |: :   ||     |  ||||||||||||||||||||||||||||||||||||
Db          8 LTKNHGYDHS--SLWRSTALEPYNLLFVINSIKQEIVNRVQNPRDERGPNMGQKLEILIK 65

Qy        599 DTL 601
              |||
Db         66 DTL 68


RESULT 5
US-09-894-273-2
; Sequence 2, Application US/09894273
; Publication No. US20040037847A1
; GENERAL INFORMATION:
;  APPLICANT: Kieff, Elliott D.
;  APPLICANT:  Ballestas, Mary E.
;  APPLICANT:  Kaye, Kenneth M.
;  TITLE OF INVENTION: RHADINO VIRUS LANA ACTS IN TRANS ON A UNIT OF RHADINO
;  TITLE OF INVENTION:  VIRUS DNA TO MEDIATE EFFICIENT EPISOME PERSISTENCE
;  FILE REFERENCE: 16412-10001R
;  CURRENT APPLICATION NUMBER: US/09/894,273
;  CURRENT FILING DATE:  2001-06-28
;  PRIOR APPLICATION NUMBER: US 60/109,422
;  PRIOR FILING DATE: 1998-11-19
;  NUMBER OF SEQ ID NOS: 3
;  SOFTWARE: PatentIn Ver. 2.0
; SEQ ID NO 2
;   LENGTH: 1162
;   TYPE: PRT
;   ORGANISM: Kaposi's sarcoma-associated herpesvirus
US-09-894-273-2

  Query Match             5.9%;  Score 196.5;  DB 3;  Length 1162;
  Best Local Similarity   24.2%;  Pred. No. 0.00056;
  Matches   96;  Conservative   52;  Mismatches  145;  Indels  103;  Gaps   14;

Qy        124 SSPYQDEEVISKTPTLAQLNSEDSQSVSDSLYYPDSLFSVKQNPLPSSFPGKKITSRAAA 183
              |||      :|   |       |: | | :|    |  | :: || ||  |:  :|
Db        111 SSPIPPSHPVSPGTT-------DTHSPSPALPPTQSPESSQRPPL-SSPTGRPDSSTPMR 162

Qy        184 PVCSSKTLQAEVPLSDCVQKASK------PPSSTQIMVKTNMYHNEKVNFHVECKDYVKK 237
              |  | :|     | :   :  ||       ||: : : |  :                :
Db        163 PPPSQQTTPPHSPTTPPPEPPSKSSPDSLAPSTLRSLRKRRLSS-------------PQ 208

Qy        238 AKVKINPVQQSRPLL--------SQIHTDAAKENTCYCGAVAKRQEKKGMEPLQ----GH 285
              :||: || |:           ::   | |:  | |: :           |      |
Db        209 GPSTLNPICQSPPVSPPRCDFANRSVYPPWATESPIYVGSSSDGDTPPRQPPTSPISIGS 268

Qy        286 ATPALPF--KETQELLL--------SPLPQEGPGSLAAGESSSLSASTSVSDSSQKKEEH 335
              ::|:      :|  |:|        |   :|   :  |||    :    :|  ||  ::
Db        269 SSPSEGSWGDDTAMLVLLAEIAEEASKNEKECSENNQAGED---NGDNEISKESQVDKDD 325

Qy        336 NYSLFVSDNLGEQPTKCSPEEDEEDEE--------------------------DVDDED 368
              |      ||  ::  : : ||||||:|                          : |||:
Db        326 N------DNKDDEEEQETDEEDEEDDEEDDEEDDEEDDEEDDEEDDEEDDEEDDEEDEE 379

Qy        369 HDEGFGSEHELSENEEEEEEEEDYEDDKDDDISDTFSEPGYENDSVEDLKE--------- 419
               ||     | :  |:|||:|||||  ||| |:|  |       |:|  || ||
Db        380 DDEEDDEEEDEEEDEEEDEEEEDEDDDDEDNEDE------EDDEEEDKKEDEEDGGDGN 433

Qy        420 -VTSISSRKRGKRRYFWEYSEQLTPSQQERMLRPSE 454
               || | :: :    |  :|  | ||| :  | :
Db        434 KTLSIQSSQQQQEPQQQE-PQQQEPQQQEPLQEPQQ 468


RESULT 6
US-10-294-804-2
; Sequence 2, Application US/10294804
; Publication No. US20030133948A1
; GENERAL INFORMATION:
;  APPLICANT: Robertson, Erle S.
;  APPLICANT:  Cotter, Murray A.
;  TITLE OF INVENTION: Methods to Inhibit or Enhance the Binding of Viral DNA
;  TITLE OF INVENTION:  to Genomic Host DNA
;  FILE REFERENCE: UM-03778
;  CURRENT APPLICATION NUMBER: US/10/294,804
;  CURRENT FILING DATE:  2002-11-14
;  PRIOR APPLICATION NUMBER: US/09/410,399
;  PRIOR FILING DATE: 1999-10-01
;  NUMBER OF SEQ ID NOS: 6
;  SOFTWARE: PatentIn Ver. 2.0
; SEQ ID NO 2
;   LENGTH: 1162
;   TYPE: PRT
;   ORGANISM: Kaposi's sarcoma-associated herpesvirus
US-10-294-804-2

  Query Match             5.9%;  Score 196.5;  DB 4;  Length 1162;
  Best Local Similarity   24.2%;  Pred. No. 0.00056;
  Matches   96;  Conservative   52;  Mismatches  145;  Indels  103;  Gaps   14;

Qy        124 SSPYQDEEVISKTPTLAQLNSEDSQSVSDSLYYPDSLFSVKQNPLPSSFPGKKITSRAAA 183
              |||      :|   |       |: | | :|    |  | :: || ||  |:  :|
Db        111 SSPIPPSHPVSPGTT-------DTHSPSPALPPTQSPESSQRPPL-SSPTGRPDSSTPMR 162

Qy        184 PVCSSKTLQAEVPLSDCVQKASK------PPSSTQIMVKTNMYHNEKVNFHVECKDYVKK 237
              |  | :|     | :   :  ||       ||: : : |  :                :
Db        163 PPPSQQTTPPHSPTTPPPEPPSKSSPDSLAPSTLRSLRKRRLSS-------------PQ 208

Qy        238 AKVKINPVQQSRPLL--------SQIHTDAAKENTCYCGAVAKRQEKKGMEPLQ----GH 285
              :||: || |:           ::   | |:  | |: :           |      |
Db        209 GPSTLNPICQSPPVSPPRCDFANRSVYPPWATESPIYVGSSSDGDTPPRQPPTSPISIGS 268

Qy        286 ATPALPF--KETQELLL--------SPLPQEGPGSLAAGESSSLSASTSVSDSSQKKEEH 335
              ::|:      :|  |:|        |   :|   :  |||    :    :|  ||  ::
Db        269 SSPSEGSWGDDTAMLVLLAEIAEEASKNEKECSENNQAGED---NGDNEISKESQVDKDD 325

Qy        336 NYSLFVSDNLGEQPTKCSPEEDEEDEE--------------------------DVDDED 368
              |      ||  ::  : : ||||||:|                          : |||:
Db        326 N------DNKDDEEEQETDEEDEEDDEEDDEEDDEEDDEEDDEEDDEEDDEEDDEEDEE 379

Qy        369 HDEGFGSEHELSENEEEEEEEEDYEDDKDDDISDTFSEPGYENDSVEDLKE--------- 419
               ||     | :  |:|||:|||||  ||| |:|  |       |:|  || ||
Db        380 DDEEDDEEEDEEEDEEEDEEEDEDDDDEDNEDE------EDDEEEDKKEDEEDGGDGN 433

Qy        420 -VTSISSRKRGKRRYFWEYSEQLTPSQQERMLRPSE 454
               || | :: :    |  :|  | ||| :  | :
Db        434 KTLSIQSSQQQQEPQQQE-PQQQEPQQQEPLQEPQQ 468
```

```
RESULT 7
US-10-194-046-2
; Sequence 2, Application US/10194046
; Publication No. US20040248081A1
; GENERAL INFORMATION:
;  APPLICANT: Robertson, Erle S.
;  TITLE OF INVENTION: LANA Viral Protein Binding Sites
;  FILE REFERENCE: UM-07142
;  CURRENT APPLICATION NUMBER: US/10/194,046
;  CURRENT FILING DATE:  2002-07-11
;  PRIOR APPLICATION NUMBER: 09/410,399
;  PRIOR FILING DATE: 1999-10-01
;  NUMBER OF SEQ ID NOS: 22
;  SOFTWARE: PatentIn version 3.1
; SEQ ID NO 2
;   LENGTH: 1162
;   TYPE: PRT
;   ORGANISM: Kaposi's sarcoma-associated herpesvirus
US-10-194-046-2

  Query Match             5.9%;  Score 196.5;  DB 5;  Length 1162;
  Best Local Similarity   24.2%;  Pred. No. 0.00056;
  Matches   96;  Conservative   52;  Mismatches  145;  Indels  103;  Gaps   14;

Qy        124 SSPYQDEEVISKTPTLAQLNSEDSQSVSDSLYYPDSLFSVKQNPLPSSFPGKKITSRAAA 183
              |||       :|   |        |: | | :|     |  | :: || ||  |:  :|
Db        111 SSPIPPSHPVSPGTT-------DTHSPSPALPPTQSPESSQRPPL-SSPTGRPDSSTPMR 162

Qy        184 PVCSSKTLQAEVPLSDCVQKASK------PPSSTQIMVKTNMYHNEKVNFHVECKDYVKK 237
              |  | :|     | :   :  ||       ||: : : |  :                 :
Db        163 PPPSQQTTPPHSPTTPPPEPPSKSSPDSLAPSTLRSLRKRRLSS--------------PQ 208

Qy        238 AKVKINPVQQSRPLL--------SQIHTDAAKENTCYCGAVAKRQEKKGMEPLQ----GH 285
                :||: || |:           ::   | |:  | |: :          |       |
Db        209 GPSTLNPICQSPPVSPPRCDFANRSVYPPWATESPIYVGSSSDGDTPPRQPPTSPISIGS 268

Qy        286 ATPALPF--KETQELLL--------SPLPQEGPGSLAAGESSSLSASTSVSDSSQKKEEH 335
              ::|:      :|  |:|        |   :|   :  |||    :    :|  ||  ::
Db        269 SSPSEGSWGDDTAMLVLLAEIAEEASKNEKECSENNQAGED---NGDNEISKESQVDKDD 325

Qy        336 NYSLFVSDNLGEQPTKCSPEEDEEDEE--------------------------DVDDED 368
              |      ||  ::  : : ||||||:|                         : |||:
Db        326 N------DNKDDEEEQETDEEDEEDDEEDDEEDDEEDDEEDDEEDDEEDDEEDDEEDEE 379

Qy        369 HDEGFGSEHELSENEEEEEEEEDYEDDKDDDISDTFSEPGYENDSVEDLKE--------- 419
              ||    | :  |:|||:||||| ||| |:|  |       |:|  || ||
Db        380 DDEEDDEEEDEEEDEEEDEEEEDEEDDDDEDNEDE------EDDEEEDKKEDEEDGGDGN 433

Qy        420 -VTSISSRKRGKRRYFWEYSEQLTPSQQERMLRPSE 454
               || | :: :     |  :|  | ||| :  | :
Db        434 KTLSIQSSQQQQEPQQQE-PQQQEPQQQEPLQEPQQ 468


RESULT 8
US-10-732-923-3319
; Sequence 3319, Application US/10732923
; Publication No. US20050108791A1
; GENERAL INFORMATION:
;  APPLICANT: Edgerton, Michael D
;  TITLE OF INVENTION: TRANSGENIC PLANTS WITH IMPROVED PHENOTYPES
;  FILE REFERENCE: 38-15(52796)C
;  CURRENT APPLICATION NUMBER: US/10/732,923
;  CURRENT FILING DATE:  2003-12-10
;  PRIOR APPLICATION NUMBER: 10/310,154
;  PRIOR FILING DATE: 2002-12-04
;  NUMBER OF SEQ ID NOS: 24149
; SEQ ID NO 3319
;   LENGTH: 1003
;   TYPE: PRT
;   ORGANISM: Kaposi's sarcoma-associated herpesvirus
US-10-732-923-3319

  Query Match             5.8%;  Score 194.5;  DB 5;  Length 1003;
  Best Local Similarity   24.5%;  Pred. No. 0.00063;
  Matches   94;  Conservative   55;  Mismatches  148;  Indels   87;  Gaps   14;

Qy        124 SSPYQDEEVISKTPTLAQLNSEDSQSVSDSLYYPDSLFSVKQNPLPSSFPGKKITSRAAA 183
              |||       :|   |        |: | | :|     |  | :: || ||  |:  :|
Db        111 SSPIPPSHPVSPGTT-------DTHSPSPALPPTQSPESSQRPPL-SSPTGRPDSSTPMR 162

Qy        184 PVCSSKTLQAEVPLSDCVQKASK------PPSSTQIMVKTNMYHNEKVNFHVECKDYVKK 237
              |  | :|     | :   :  ||       ||: : : |  :                 :
Db        163 PPPSQQTTPPHSPTTPPPEPPSKSSPDSLAPSTLRSLRKRRLSS--------------PQ 208

Qy        238 AKVKINPVQQSRPLL--------SQIHTDAAKENTCYCGAVAKRQEKKGMEPLQ----GH 285
                :||: || |:           ::   | |:  | |: :          |       |
Db        209 GPSTLNPICQSPPVSPPRCDFANRSVYPPWATESPIYVGSSSDGDTPPRQPPTSPISIGS 268

Qy        286 ATPALPF--KETQELLL--------SPLPQEGPGSLAAGESSSLSASTSVSDSSQKKEEH 335
              ::|:      :|  |:|        |   :|   :  |||    :    :|  ||  ::
Db        269 SSPSEGSWGDDTAMLVLLAEIAEEASKNEKECSENNQAGED---NGDNEISKESQVDKDD 325

Qy        336 NYSLFVSDNLGEQPTKCSPEEDEEDEE--------------DVDDEDHDEGFGSEHELS 380
              |      ||  ::  : : ||||||:|               : |||: ||    | :
Db        326 N------DNKDDEEEQETDEEDEEDDEEDDEEDDEEDDEEDDEEDDEEDDEEDDEEDEE 379

Qy        381 ENEEEEEEEEDYEDDKDDDISDTFSEPGYENDSVEDLKE------VTSISSRKRGKRRYF 434
              |:||||:|||: ||::|||  |   |  | |  || ::         || | :: :
Db        380 EDEEEEDEEEEEEDEEDDDDEDNEDE---EEDKKEDEEDGGDGNKTLSIQSSQQQQEPQQ 436

Qy        435 WE----YSEQLTPSQQERMLRPSE 454
               |      :|  | ||| :  | :
Db        437 QEPQQQEPQQQEPQQQEPLQEPQQ 460


RESULT 9
US-10-483-505-3
; Sequence 3, Application US/10483505
; Publication No. US20050118582A1
; GENERAL INFORMATION:
;  APPLICANT: INCYTE CORPORATION; SWARNAKAR, Anita;
;  APPLICANT:  RICHARDSON, Thomas W.; WARREN, Bridget A.;
;  APPLICANT:  GRIFFIN, Jennifer A.; TANG, Y. Tom;
;  APPLICANT:  YUE, Henry; BAUGHN, Mariah R.;
;  APPLICANT:  EMERLING, Brooke M.; LAL, Preeti G.;
;  APPLICANT:  LU, Dyung Aina M.; FORSYTHE, Ian J.;
;  APPLICANT:  RAMKUMAR, Jayalaxmi; LI, Joana X.;
;  APPLICANT:  BECHA, Shanya D.; DUGGAN, Brendan M.;
;  APPLICANT:  SANJANWALA, Madhusudan M.; LEE, Ernestine A.;
;  APPLICANT:  BURFORD, Neil; ELLIOTT, Vicki S.;
;  APPLICANT:  ISON, Craig H.; DING, Li;
;  APPLICANT:  BOROWSKY, Mark L.; YAO, Monique G.;
;  APPLICANT:  BARROSO, Ines; TRAN, Bao;
;  APPLICANT:  WALIA, Narinder K.; HAFALIA, April J. A.;
;  APPLICANT:  NGUYEN, Danniel B.; LU, Yan;
;  APPLICANT:  ARVIZU, Chandra S.
;  TITLE OF INVENTION: NUCLEIC ACID-ASSOCIATED PROTEINS
;  FILE REFERENCE: PF-1068 USN
;  CURRENT APPLICATION NUMBER: US/10/483,505
;  CURRENT FILING DATE:  2004-12-12
;  PRIOR APPLICATION NUMBER: PCT/US02/21971
;  PRIOR FILING DATE: 2002-07-10
;  PRIOR APPLICATION NUMBER: US 60/305,089
;  PRIOR FILING DATE: 2001-07-12
;  PRIOR APPLICATION NUMBER: US 60/305,104
;  PRIOR FILING DATE: 2001-07-12
;  PRIOR APPLICATION NUMBER: US 60/305,325
;  PRIOR FILING DATE: 2001-07-13
;  PRIOR APPLICATION NUMBER: US 60/305,390
;  PRIOR FILING DATE: 2001-07-13
;  PRIOR APPLICATION NUMBER: US 60/306,960
;  PRIOR FILING DATE: 2001-07-19
;  PRIOR APPLICATION NUMBER: US 60/306,694
;  PRIOR FILING DATE: 2001-07-20
```

```
;  PRIOR APPLICATION NUMBER: US 60/308,170
;  PRIOR FILING DATE: 2001-07-27
;  NUMBER OF SEQ ID NOS: 70
;  SOFTWARE: PERL Program
; SEQ ID NO 3
;    LENGTH: 735
;    TYPE: PRT
;    ORGANISM: Homo sapiens
;    FEATURE:
;    NAME/KEY: misc_feature
;    OTHER INFORMATION: Incyte ID No: 7037554CD1
US-10-483-505-3

  Query Match             5.7%;  Score 189.5;  DB 5;  Length 735;
  Best Local Similarity   23.1%;  Pred. No. 0.00089;
  Matches  134;  Conservative   78;  Mismatches  199;  Indels  169;  Gaps   28;

Qy        120 LDDFSS--PYQDEEVISKTPTLAQLNSEDSQSVSDSLYYPDSLFSVKQNPLPSSFPGKKI 177
              |||  :  | ||:|: :      :   : |:  || :   :|  : :| |   |   :::
Db         16 LDDILTEVPEQDDELYNPESEQDKNEKKGSKRKSDRM---ESTDTKRQKP---SVHSRQL 69

Qy        178 TSRAAAPVCSSKTLQAEVPLSDCVQK-----ASKPPSSTQIMVKTNMYHNEKVNFHVECK 232
               |:              |||  |       ::|  |:|:       | ||:       |
Db         70 VSK--------------PLSSSVSNNKRIVSTKGKSATE-------YKNEEYQRSERNK 107

Qy        233 DYVKKAKVKINPVQQSRPLLSQIHTDAAKENTCYCGAVAKRQEKKGMEPLQGHATPALPF 292
                  |::::     |  :|     | ||    | ||  ::  :
Db        108 RLDADRKIRLSSSASREPYKNQ------PEKTC----VRKRDPERRAK----------- 145

Qy        293 KETQELLLSPLPQEGPGSLAAGESSSLSASTSVSDSSQKKEEHNY-------SLFVSDNL 345
                  || |    ||   |      || |   | ::   |       |   ||
Db        146 --------SPTPD---GSERIGLEVDRRASRSSQSSKEEVNSEEYGSDHETGSSGSSDEQ 194

Qy        346 G------EQPTKCSPEEDEEDEEDVD-----DEDHDEGFGSEHELSENEEEEEEEEDYED 394
              |      |:  :   ||||| ||| :     ||| :|    | |  | |||||||||:||
Db        195 GNNTENEEEGVEEDVEEDEEVEEDAEEDEEVDEDGEE--EEEEEEEEEEEEEEEEEYEQ 252

Qy        395 DKDD--------DISDTFSEPGYENDSVED----LKEVTSISSRKRGKRRYFWEYSEQLT 442
              |: |        |    |: | |: | |        |  |  |: :|:
Db        253 DERDQKEEGNDYDTRSEASDSGSESVSFTDGSVRSGSGTDGSDEKKKERK---------- 302

Qy        443 PSQQERMLRPSEWNRD-TLPSNMYQKNGLHHGKY-----AVKKSRRTDVEDLTPNPKKLL 496
                : | : |  ::|  :  |  |  :   | |      ||:| : : :: :  : :  |
Db        303 ---RARGISPIVFDRSGSSASESYAGSEKKHEKLSSSVRAVRKDQTSKLKYVLQDARFFL 359

Qy        497 QIGN-----ELRKLNKVISDLTPVSELPLTARPRSRKEKNKLAFRA-------------- 537
                  |     | |   | | | ||:|          :|  ||||:
Db        360 IKSNNHENVSLAKAKGVWSTL-PVNE----------KKLNLAFRSARSVILIFSVRESG 407

Qy        538 -----CRLKKKAQYEANKVKLW----GLNTEYDNLLFVINSI-KQEI----VNRVQNPRD 583
                  ||  :: :  : :  |    |:: :    :| |: | ::|:      : || :
Db        408 KFQGFARLSSESHHGGSPIH-WVLPAGMSAKMLGGVFKIDWICRRELPFTKSAHLTNPWN 466

Qy        584 ERGP----NMGQKLEILIKDTLGLPVAGQTSEFVNQVLEK 619
              |  |      ||::|:     | |       |  : ||: |
Db        467 EHKPVKIGRDGQEIELECGTQLCLLFPPDESIDLYQVIHK 506


RESULT 10
US-10-032-585-7386
; Sequence 7386, Application US/10032585
; Publication No. US20030180953A1
; GENERAL INFORMATION:
;  APPLICANT: Terry, Roemer D.
;  APPLICANT:  Bo, Jiang
;  APPLICANT:  Charles, Boone
;  APPLICANT:  Howard, Bussey
;  TITLE OF INVENTION: Gene Disruption Methodologies for Drug Target Discovery
;  FILE REFERENCE: 10182-005-999
;  CURRENT APPLICATION NUMBER: US/10/032,585
;  CURRENT FILING DATE:  2001-12-20
;  NUMBER OF SEQ ID NOS: 8000
;  SOFTWARE: PatentIn version 3.1

; SEQ ID NO 7386
;    LENGTH: 427
;    TYPE: PRT
;    ORGANISM: Candida albicans
US-10-032-585-7386

  Query Match             5.3%;  Score 179;  DB 4;  Length 427;
  Best Local Similarity   22.2%;  Pred. No. 0.0021;
  Matches  106;  Conservative   69;  Mismatches  150;  Indels  152;  Gaps   20;

Qy        197 LSDCVQKA-SKPPSSTQIMVKTNMYHNEKVNFHVECKDYVKKAKV--KINPVQQSRPLLS 253
              | | : || |||   | :  ||            | |: :: | |  |:   ::|:
Db         15 LIDAIDKAKSKPTKKTPVTEKT----------VEEEKEIIEAAPVTTKVKSNKKSK---- 60

Qy        254 QIHTDAAKENTCYCGAVAKRQEKKGMEPLQGHATPALPFKETQELLLSPLPQEGPGSLAA 313
                  ::  |::  |   |  ::|:: :: ||          |: :|
Db         61 --KSNTKKDDDDYQSEVLSKKEQRRLKKLQ--------TKQQEE--------------- 94

Qy        314 GESSSLSASTSVSDSSQKKEEHNYSLFVSDNLGEQPTKCSPEEDEEDEEDVDDE--DHDE 371
                            :|: |:|           ||:|||::| |:|  | ::
Db         95 ---------------EKEAENN---------------EEEEEEDDDDDDEEELDLEK 121

Qy        372 GFGSEHELSENEEEEE------EEEDYEDDKDDDISDTFSEPGYENDSVEDLKEVTSISS 425
                || |   |:||||      |||| | : ||:  |       | ||  |:   | ::
Db        122 LAASESESDINDEEEEDDDDDDEEEDDEVENDDEAEDDIPLSDVEVDSDADIVPHTKLTI 181

Qy        426 RKRG------------KRRYFWEYSEQLTPSQQERMLRPSEWNRDTLPSNMYQKNGLHH 472
                              :  | |:    :  : |  ::  :   ||     : | ||
Db        182 NNMAALRESLARIELPWSKHSFIEHQSITSADKTESEIK--DIYDDTERELAFYKQGLD- 238

Qy        473 GKYAVKKSRRTDVEDLTPNPKKLLQIGNELRKLNKVISDLTPVSELPLTARPRSRKEKNK 532
                |||:||:|           ||::    :     :::      |          | |||
Db        239 ---AVKQSRKT----------LLKLKIPFSRPMDYFAEMVKSDE-------HMDKLKNK 277

Qy        533 LAFRACRLKKKAQYEANK---VKLWGLNTEYDNLLFVINSIKQEIVNRVQNPRDERGPNM 589
              |   |    |||  || :   :| :|   ::  |        |:| : :::: : :|| |
Db        278 LLTEAA--NKKASEEAKRQRQLKKFGKQVQHATLQERAKQ-KKETLEKIKSLKKKRGAN- 333

Qy        590 GQKLEILIKDTLGLPVAGQTSEFVNQVLEKTAE--------GNPTGGLVGLRIPTSK 638
                ||  |   :                ||:  |       |: :||    | | ||
Db        334 ----EISNDDDFQI------------ALEEATENNQKYGHGGSGSGGDNKRRKPNSK 374


RESULT 11
US-10-487-593-1
; Sequence 1, Application US/10487593
; Publication No. US20050064413A1
; GENERAL INFORMATION:
;  APPLICANT: NAKAGAWARA, AKIRA
;  APPLICANT:  HATTORI, MASAHIRA
;  APPLICANT:  SAKAKI, YOSHIYUKI
;  TITLE OF INVENTION: NOVEL HUMAN BMCC1 GENE
;  FILE REFERENCE: 7388/80977
;  CURRENT APPLICATION NUMBER: US/10/487,593
;  CURRENT FILING DATE:  2004-02-24
;  PRIOR APPLICATION NUMBER: PCT/JP02/08520
;  PRIOR FILING DATE: 2002-08-23
;  PRIOR APPLICATION NUMBER: JP 2001-255198
;  PRIOR FILING DATE: 2001-08-24
;  NUMBER OF SEQ ID NOS: 55
;  SOFTWARE: PatentIn Ver. 3.2
; SEQ ID NO 1
;    LENGTH: 2724
;    TYPE: PRT
;    ORGANISM: Homo sapiens
US-10-487-593-1

  Query Match             5.2%;  Score 175;  DB 5;  Length 2724;
  Best Local Similarity   19.5%;  Pred. No. 0.044;
  Matches  165;  Conservative   87;  Mismatches  288;  Indels  308;  Gaps   38;

Qy          2 PQPSVSGMD---------PPFGDAFRSHTFSEQTL---------MSTDL-------LANS 36
              | |  ||           |  | | :   ||  |         | |::       | :
```

```
Db        139 PAPMASGQSQQSSHSADYSPADDFFPNSDLSEGQLPAGPEGLDGMGTNMSNYSSSSLLSG 198

Qy         37 SDPDFMYELDRE-MNYQQNPRDNF---LSLEDCKDIENLESFTDVLDNEG---------- 82

Db        199 AGKDSLVEHDEEFVQRQDSPRDNSERNLSLTDFVGDES--PSPERLKNTGKRIPPTPMNS 256

Qy         83 ---ALTSNWEQWDTYCEDLTK----------YTKLTSCDIWGTKEVDYLGLDDFSSPYQ 128

Db        257 LVESSPSTEEPASLYTEDMTQKATDTGHMGPPQTHARCSSWWGGLEIDSKNIADAWSSSE 316

Qy        129 DEEVISKTPTLAQLN-SEDSQSVSDSLYYPDSLFSVKQNPLPSSFPGKKITSRAAAPVCS 187

Db        317 QESVFQSPESWKEHKPSSIDRRASDSVFQPKSLEFTKSGPWESEF---------GQPELG 367

Qy        188 SKTLQ----AEVPLSDC-VQKASKP---PSSTQIMVK--TNMYHNEK------------ 224

Db        368 SNDIQDKNEESLPFQNLPMEKSPLPNTSPQGTNHLIEDFASLWHSGRSPTAMPEPWGNPT 427

Qy        225 -------------VNFHVECKDYVKKAKVKINPVQQSRPLL------------------ 252

Db        428 DDGEPAAVAPFPAWSAFGKEDHDEALKNTWNLHPTSSKTPSVRDPNEWAMAKSGFAFSSS 487

Qy        253 -------SQIHTDAA------KENTCYCGAVAKRQEKKGMEPLQGHA-TPALPFKETQEL 298

Db        488 ELLDNSPSEINNEAAPEIWGKKNNDSRDHIFAPGNPSSDLD----HTWTNSKPPKEDQNG 543

Qy        299 LLSP------------------LPQEGPGSLAAGESSSL-------SASTSVSDSSQKK 332

Db        544 LVDPKTRGKVYEKVDSWNLFEENMKKGGSDVLVPWEDSFLSYKCSDYSASNLGEDSVPSP 603

Qy        333 EEHNYS--------LFVSDNLGEQPTKCSPEEDEE-----------DEEDVDDE------ 367

Db        604 LDTNYSTSDSYTSPTFAGD---EKETEHKPFAKEEGFESKDGNSTAEETDIPPQSLQQSS 660

Qy        368 -------------------DHDEGFGSEHELSENEEEEEEEEDYEDDKDDDISDTFSEPG 408

Db        661 RNRISSGPGNLDMWASPHTDNSSEINTTHNLDENELKTEHTDGKNISMEDDVGES-SQSS 719

Qy        409 YENDSVEDLKEVTSISSRKRGKRRYFWEYSEQLT---PSQQERMLRPSE---WNR----D 458

Db        720 YDDPSMMQL----------------YNETNRQLTLLHSSTNSRQTAPDSLDLWNRVILED 763

Qy        459 T----------------------LPSNMYQKNGLHHGKYAVKKSRRTDVEDLTPNPKKL 495

Db        764 TQSTATISDMDNDLDWDDCSGGAAIPSDGQTEGYMAEGS---EPETRFTVRQLEP----- 815

Qy        496 LQIGNELRKLNKVISDLTPVSELPLTARPRSRKEKNKLAFRACRLKKKAQYEANKVKLW- 554

Db        816 --WGLEYQEANQVDWELPASDEHTKDSAPSEHHTLN---------EKSGQLIANSI--WD 862

Qy        555 GLNTEYDNLLFVI-----------NSIKQEIVNRVQNPRDERGPNMGQKLEILIKDTLGL 603

Db        863 SVMRDKDMSSFMLPGSSHITDSEQRELPPEIPSHSANVKDTHSPD-------------A 908

Qy        604 PVAGQTSE 611

Db        909 PAASGTSE 916


RESULT 12
US-09-825-886-22
; Sequence 22, Application US/09825886
; Publication No. US20020076693A1
; GENERAL INFORMATION:
;  APPLICANT: Hovanessian, Ara
;  APPLICANT:  Callebaut, Christian
;  APPLICANT:  Krust, Bernard
;  APPLICANT:  Jacotot, Etienne
;  APPLICANT:  Muller, Sylviane
;  APPLICANT:  Briand, Jean-Paul
;  APPLICANT:  Guichard, Giles
;  TITLE OF INVENTION: A NOVEL CELL SURFACE RECEPTOR FOR HIV RETROVIRUSES,
;  TITLE OF INVENTION:  THERAPEUTIC AND DIAGNOSTIC USES.
;  FILE REFERENCE: 03495.0166-01000
;  CURRENT APPLICATION NUMBER: US/09/825,886
;  CURRENT FILING DATE:  2001-07-26
;  PRIOR APPLICATION NUMBER: 09/393,302
;  PRIOR FILING DATE: 1999-09-10
;  PRIOR APPLICATION NUMBER: PCT/EP98/01409
;  PRIOR FILING DATE: 1998-03-12
;  PRIOR APPLICATION NUMBER: 60/040,969
;  PRIOR FILING DATE: 1997-03-12
;  NUMBER OF SEQ ID NOS: 32
;  SOFTWARE: PatentIn Ver. 2.1
; SEQ ID NO 22
;   LENGTH: 707
;   TYPE: PRT
;   ORGANISM: Homo sapiens
US-09-825-886-22

  Query Match             5.2%;  Score 174;  DB 3;  Length 707;
  Best Local Similarity   20.4%;  Pred. No. 0.0088;
  Matches   99;  Conservative   76;  Mismatches  162;  Indels  148;  Gaps   20;

Qy        129 DEEVISKTPTLAQLNSEDSQSVSDSLYYPDSLFSVKQNPLPSSFPGKKITSRAAAPVCSS 188

Db         13 DPKKMAPPPKEVEEDSEDEEMSEDE--EDDS--SGEEVVIPQK-KGKKAAATSAKKVVVS 67

Qy        189 KTLQAEVPLSDCVQKASKPPSSTQIMVKTNMYHNEKVNFHVECKDYVKKAKVKINPVQQS 248

Db         68 PT--KKVAVATPAKKAAVTPGK--------------------KAAATPAKKTVTPAK-- 102

Qy        249 RPLLSQIHTDAAKENTCYCGAVAKRQEKKGMEPLQGHATPALPFKETQELLLSPLPQEGP 308

Db        103 ---------------------AVTTPGKKGATP--GKALVATPGKKG-----AAIPAKGA 134

Qy        309 GSLAAGESSSLSASTSVSDSSQKKEEHNYSLFVSDNLGEQP--------TKCSPEEDEED 360

Db        135 KN---GKNAKKEDSDEEEDDDSEEDEEDDEDEDEDEDEIEPAAMKAAAAAPASEDEDDED 191

Qy        361 EEDVDDEDHDEGFGSEHE---------------------LSENEEEEEEEEDYEDDKDD 398

Db        192 DEDDEDDDDDEEDDSEEEAMETTPAKGKKAAKVVPVKAKNVAEDEDEEEDDEDEDDDDDE 251

Qy        399 DISDTFSEPGYENDSVEDLKEVTSISSRKRGKRRYFWEYSEQLTPSQQERMLRPSE---- 454

Db        252 DDEDDDDEDDEEEEEEEEEPVKEAPGKRKKEMA-----KQKAAPEAKKQKVEGTEPTTA 306

Qy        455 ---------WNRD-----TLPSNMYQKNGLHHGKYAVKKSRR------TDVEDLTPNPKK 494

Db        307 FNLFVGNLNFNKSAPELKTGISDVFAKNDLAVVDVRIGMTRKFGYVDFESAEDL----EK 362

Qy        495 LLQI------GNELRKLNKVISDLTPVSELPLTARPRSRKEKNKLAFRACRLKKKAQYEA 548

Db        363 ALELTGLKVFGNEI-KLEK---------------PKGKDSKKERDARTL-LAKNLPYKV 404

Qy        549 NKVKL 553

Db        405 TQDEL 409


RESULT 13
US-10-384-569-3
; Sequence 3, Application US/10384569
; Publication No. US20040002457A1
; GENERAL INFORMATION:
;  APPLICANT: HOVANESSIAN, ARA G.
;  APPLICANT:  BRIAND, JEAN-PAUL
;  TITLE OF INVENTION: INHIBITION OF VIRUS ANCHORAGE BY RGG DOMAIN OF A CELL SURFACE-
;  TITLE OF INVENTION:  EXPRESSED PROTEIN, POLYNUCLEOTIDE CODING FOR SAID RGG DOMAIN,
;  TITLE OF INVENTION:  THERAPEUTIC USES THEREOF BY INHIBITION OF MICROORGANISM OR
;  TITLE OF INVENTION:  PROTEIN LIGAND BINDING TO THE CELL-SURFACE-EXPRESSED PROTEIN
;  FILE REFERENCE: 03495.0260-00000
;  CURRENT APPLICATION NUMBER: US/10/384,569
;  CURRENT FILING DATE:  2004-07-14
;  PRIOR APPLICATION NUMBER: 60/363,371
```

```
;  PRIOR FILING DATE: 2002-03-12
;  PRIOR APPLICATION NUMBER: 60/397,600
;  PRIOR FILING DATE: 2002-07-23
;  NUMBER OF SEQ ID NOS: 25
;  SOFTWARE: PatentIn Ver. 2.1
; SEQ ID NO 3
;   LENGTH: 707
;   TYPE: PRT
;   ORGANISM: Homo sapiens
US-10-384-569-3

  Query Match             5.2%;  Score 174;  DB 4;  Length 707;
  Best Local Similarity   20.4%;  Pred. No. 0.0088;
  Matches   99;  Conservative   76;  Mismatches  162;  Indels  148;  Gaps   20;

Qy        129 DEEVISKTPTLAQLNSEDSQSVSDSLYYPDSLFSVKQNPLPSSFPGKKITSRAAAPVCSS 188
              | : ::  |   : :|||  :   |     ||  | ::  :|    |||  : :|  |  |
Db         13 DPKKMAPPPKEVEEDSEDEEMSEDE--EDDS--SGEEVVIPQK-KGKKAAATSAKKVVVS 67

Qy        189 KTLQAEVPLSDCVQKASKPPSSTQIMVKTNMYHNEKVNFHVECKDYVKKAKVKINPVQQS 248
              |   :| ::   :||:  |                          |     ||  : | :
Db         68 PT--KKVAVATPAKKAAVTPGK--------------------KAAATPAKKTVTPAK-- 102

Qy        249 RPLLSQIHTDAAKENTCYCGAVAKRQEKKGMEPLQGHATPALPFKETQELLLSPLPQEGP 308
                                   |||  |  | |  | | |:      : :| :|
Db        103 --------------------AVTTPGKKGATP--GKALVATPGKKG-----AAIPAKGA 134

Qy        309 GSLAAGESSSLSASTSVSDSSQKKEEHNYSLFVSDNLGEQP--------TKCSPEEDEED 360
              :   |:::    |    |   :::| :      |    :|           | :||:||
Db        135 KN---GKNAKKEDSDEEEDDDSEEDEEDDEDEDEDEDEIEPAAMKAAAAAPASEDEDDED 191

Qy        361 EEDVDDEDHDEGFGSEHE--------------------LSENEEEEEEEEDYEDDKDD 398
              :|| :|:| ||   || |                    ::|:|:|||::|| :|| |:
Db        192 DEDDEDDDDDEEDDSEEEAMETTPAKGKKAAKVVPVKAKNVAEDEDEEEDDEDEDDDDDE 251

Qy        399 DISDTFSEPGYENDSVEDLKEVTSISSRKRGKRRYFWEYSEQLTPSQQERMLRPSE---- 454
              |  |   |   | :  |: : |     ::: :        ::  |  ::: :  :|
Db        252 DDEDDDDEDDEEEEEEEEEPVKEAPGKRKKEMA-----KQKAAPEAKKQKVEGTEPTTA 306

Qy        455 ---------WNRD-----TLPSNMYQKNGLHHGKYAVKKSRR------TDVEDLTPNPKK 494
                      :|:      |  |::: || |      :  :|:          |||    :|
Db        307 FNLFVGNLNFNKSAPELKTGISDVFAKNDLAVVDVRIGMTRKFGYVDFESAEDL----EK 362

Qy        495 LLQI------GNELRKLNKVISDLTPVSELPLTARPRSRKEKNKLAFRACRLKKKAQYEA 548
              |::      |||: || |                |: :  | :   |  | |   |:
Db        363 ALELTGLKVFGNEI-KLEK---------------PKGKDSKKERDARTL-LAKNLPYKV 404

Qy        549 NKVKL 553
                : :|
Db        405 TQDEL 409


RESULT 14
US-10-400-083-19
; Sequence 19, Application US/10400083
; Publication No. US20040186056A1
; GENERAL INFORMATION:
;  APPLICANT: Ruoslahti, Erkki
;  APPLICANT:  Porkka, Kimmo
;  APPLICANT:  Christian, Sven
;  TITLE OF INVENTION: HMGN2 Peptides and Related Molecules
;  TITLE OF INVENTION:  that Selectively Home to Tumor Blood Vessels and Tumor Cells
;  FILE REFERENCE: P-LJ 5662
;  CURRENT APPLICATION NUMBER: US/10/400,083
;  CURRENT FILING DATE:  2003-03-20
;  PRIOR APPLICATION NUMBER: US 10/116,866
;  PRIOR FILING DATE: 2002-04-05
;  NUMBER OF SEQ ID NOS: 20
;  SOFTWARE: FastSEQ for Windows Version 4.0
; SEQ ID NO 19
;   LENGTH: 707
;   TYPE: PRT
;   ORGANISM: Homo sapiens
US-10-400-083-19

  Query Match             5.2%;  Score 174;  DB 4;  Length 707;
  Best Local Similarity   20.4%;  Pred. No. 0.0088;
  Matches   99;  Conservative   76;  Mismatches  162;  Indels  148;  Gaps   20;

Qy        129 DEEVISKTPTLAQLNSEDSQSVSDSLYYPDSLFSVKQNPLPSSFPGKKITSRAAAPVCSS 188
              | : ::  |   : :|||  :   |     ||  | ::  :|    |||  : :|  |  |
Db         13 DPKKMAPPPKEVEEDSEDEEMSEDE--EDDS--SGEEVVIPQK-KGKKAAATSAKKVVVS 67

Qy        189 KTLQAEVPLSDCVQKASKPPSSTQIMVKTNMYHNEKVNFHVECKDYVKKAKVKINPVQQS 248
              |   :| ::   :||:  |                          |     ||  : | :
Db         68 PT--KKVAVATPAKKAAVTPGK--------------------KAAATPAKKTVTPAK-- 102

Qy        249 RPLLSQIHTDAAKENTCYCGAVAKRQEKKGMEPLQGHATPALPFKETQELLLSPLPQEGP 308
                                   |||  |  | |  | | |:      : :| :|
Db        103 --------------------AVTTPGKKGATP--GKALVATPGKKG-----AAIPAKGA 134

Qy        309 GSLAAGESSSLSASTSVSDSSQKKEEHNYSLFVSDNLGEQP--------TKCSPEEDEED 360
              :   |:::    |    |   :::| :      |    :|           | :||:||
Db        135 KN---GKNAKKEDSDEEEDDDSEEDEEDDEDEDEDEDEIEPAAMKAAAAAPASEDEDDED 191

Qy        361 EEDVDDEDHDEGFGSEHE--------------------LSENEEEEEEEEDYEDDKDD 398
              :|| :|:| ||   || |                    ::|:|:|||::|| :|| |:
Db        192 DEDDEDDDDDEEDDSEEEAMETTPAKGKKAAKVVPVKAKNVAEDEDEEEDDEDEDDDDDE 251

Qy        399 DISDTFSEPGYENDSVEDLKEVTSISSRKRGKRRYFWEYSEQLTPSQQERMLRPSE---- 454
              |  |   |   | :  |: : |     ::: :        ::  |  ::: :  :|
Db        252 DDEDDDDEDDEEEEEEEEEPVKEAPGKRKKEMA-----KQKAAPEAKKQKVEGTEPTTA 306

Qy        455 ---------WNRD-----TLPSNMYQKNGLHHGKYAVKKSRR------TDVEDLTPNPKK 494
                      :|:      |  |::: || |      :  :|:          |||    :|
Db        307 FNLFVGNLNFNKSAPELKTGISDVFAKNDLAVVDVRIGMTRKFGYVDFESAEDL----EK 362

Qy        495 LLQI------GNELRKLNKVISDLTPVSELPLTARPRSRKEKNKLAFRACRLKKKAQYEA 548
              |::      |||: || |                |: :  | :   |  | |   |:
Db        363 ALELTGLKVFGNEI-KLEK---------------PKGKDSKKERDARTL-LAKNLPYKV 404

Qy        549 NKVKL 553
                : :|
Db        405 TQDEL 409


RESULT 15
US-10-491-545A-20
; Sequence 20, Application US/10491545A
; Publication No. US20050130117A1
; GENERAL INFORMATION:
;  APPLICANT: Chu, Peter
;  APPLICANT:  Li, Congfen
;  APPLICANT:  Liao, X. Charlene
;  APPLICANT:  Masuda, Esteban
;  APPLICANT:  Pardo, Jorge
;  APPLICANT:  Zhao, Haoran
;  APPLICANT:  Rigel Pharmaceuticals, Inc.
;  TITLE OF INVENTION: Modulators of Lymphocyte Activation and Migration
;  FILE REFERENCE: 021044-000330US
;  CURRENT APPLICATION NUMBER: US/10/491,545A
;  CURRENT FILING DATE:  2004-04-01
;  PRIOR APPLICATION NUMBER: US 60/327,212
;  PRIOR FILING DATE: 2001-10-03
;  PRIOR APPLICATION NUMBER: WO PCT/US02/31618
;  PRIOR FILING DATE: 2002-10-02
;  NUMBER OF SEQ ID NOS: 96
;  SOFTWARE: PatentIn Ver. 2.1
; SEQ ID NO 20
;   LENGTH: 707
;   TYPE: PRT
;   ORGANISM: Homo sapiens
;   FEATURE:
;   OTHER INFORMATION: human nucleolin (NCL)
US-10-491-545A-20
```

```
  Query Match             5.2%;  Score 174;  DB 5;  Length 707;
  Best Local Similarity   20.4%;  Pred. No. 0.0088;
  Matches   99;  Conservative   76;  Mismatches  162;  Indels  148;  Gaps   20;

Qy        129 DEEVISKTPTLAQLNSEDSQSVSDSLYYPDSLFSVKQNPLPSSFPGKKITSRAAAPVCSS 188
              | : ::  |   : :||| :   |    ||  | ::  :|    |||  : :|  |  |
Db         13 DPKKMAPPPKEVEEDSEDEEMSEDE--EDDS--SGEEVVIPQK-KGKKAAATSAKKVVVS 67

Qy        189 KTLQAEVPLSDCVQKASKPPSSTQIMVKTNMYHNEKVNFHVECKDYVKKAKVKINPVQQS 248
               |   :| ::   :||:  |                         |     ||  : | :
Db         68 PT--KKVAVATPAKKAAVTPGK--------------------KAAATPAKKTVTPAK-- 102

Qy        249 RPLLSQIHTDAAKENTCYCGAVAKRQEKKGMEPLQGHATPALPFKETQELLLSPLPQEGP 308
                                          |||  |  | |  | | |:      : :| :|
Db        103 --------------------AVTTPGKKGATP--GKALVATPGKKG-----AAIPAKGA 134

Qy        309 GSLAAGESSSLSASTSVSDSSQKKEEHNYSLFVSDNLGEQP--------TKCSPEEDEED 360
               :   |:::    |    |   :::| :      |    :|          | :||:||
Db        135 KN---GKNAKKEDSDEEEDDDSEEDEEDDEDEDEDEDEIEPAAMKAAAAAPASEDEDDED 191

Qy        361 EEDVDDEDHDEGFGSEHE----------------------LSENEEEEEEEEDYEDDKDD 398
              :|| :|:| ||   || |                      ::|:|:|:|||::|| :|| |:
Db        192 DEDDEDDDDDEEDDSEEEAMETTPAKGKKAAKVVPVKAKNVAEDEDEEEDDEDEDDDDDE 251

Qy        399 DISDTFSEPGYENDSVEDLKEVTSISSRKRGKRRYFWEYSEQLTPSQQERMLRPSE---- 454
              |  |   |   | :  |: : |     ::: :        ::  |  ::: :  :|
Db        252 DDEDDDDEDDEEEEEEEEEEPVKEAPGKRKKEMA-----KQKAAPEAKKQKVEGTEPTTA 306

Qy        455 ---------WNRD-----TLPSNMYQKNGLHHGKYAVKKSRR------TDVEDLTPNPKK 494
                      :|:       |  |::: || |       :  :|:          |||    :|
Db        307 FNLFVGNLNFNKSAPELKTGISDVFAKNDLAVVDVRIGMTRKFGYVDFESAEDL----EK 362

Qy        495 LLQI------GNELRKLNKVISDLTPVSELPLTARPRSRKEKNKLAFRACRLKKKAQYEA 548
               |::      |||: || |                 |: :  | :   |   | |   |:
Db        363 ALELTGLKVFGNEI-KLEK---------------PKGKDSKKERDARTL-LAKNLPYKV 404

Qy        549 NKVKL 553
               : :|
Db        405 TQDEL 409


Search completed: February 28, 2006, 09:02:00
Job time : 188.308 secs
```

GenCore version 5.1.7

OM protein - protein search, using sw model

Run on: February 28, 2006, 08:56:26 ; Search time 17.0827 Seconds
(without alignments)
556.876 Million cell updates/sec

Title: US-10-717-665A-44
Perfect score: 3349
Sequence: 1 MPQPSVSGMDPPFGDAFRSH.....TAEGNPTGGLVGLRIPTSKV 639

Scoring table: BLOSUM62
Gapop 10.0 , Gapext 0.5

Searched: 117670 seqs, 14887254 residues

Total number of hits satisfying chosen parameters: 117670

Minimum DB seq length: 0
Maximum DB seq length: 2000000000

Post-processing: Minimum Match 0%
Maximum Match 100%
Listing first 45 summaries

Database : Published_Applications_AA_New:*
1: /cgn2_6/ptodata/2/pubpaa/US08_NEW_PUB.pep:*
2: /cgn2_6/ptodata/2/pubpaa/US06_NEW_PUB.pep:*
3: /cgn2_6/ptodata/2/pubpaa/US07_NEW_PUB.pep:*
4: /cgn2_6/ptodata/2/pubpaa/PCT_NEW_PUB.pep:*
5: /cgn2_6/ptodata/2/pubpaa/US09_NEW_PUB.pep:*
6: /cgn2_6/ptodata/2/pubpaa/US10_NEW_PUB.pep:*
7: /cgn2_6/ptodata/2/pubpaa/US11_NEW_PUB.pep:*
8: /cgn2_6/ptodata/2/pubpaa/US60_NEW_PUB.pep:*

Pred. No. is the number of results predicted by chance to have a score greater than or equal to the score of the result being printed, and is derived by analysis of the total score distribution.

SUMMARIES

Result No.	Score	% Query Match	Length	DB	ID	Description
1	164.5	4.9	687	7	US-11-072-512-2651	Sequence 2651, Ap
2	163	4.9	354	7	US-11-189-817-2	Sequence 2, Appli
3	157.5	4.7	651	7	US-11-128-660-1	Sequence 1, Appli
4	155	4.6	1168	6	US-10-509-422-2	Sequence 2, Appli
5	154	4.6	740	7	US-11-124-367A-293	Sequence 293, App
6	154	4.6	760	7	US-11-124-367A-292	Sequence 292, App
7	147.5	4.4	1618	6	US-10-984-645-2	Sequence 2, Appli
8	145.5	4.3	795	6	US-10-770-726-49	Sequence 49, Appl
9	145.5	4.3	1735	6	US-10-495-083-3	Sequence 3, Appli
10	144	4.3	1404	6	US-10-878-556A-169	Sequence 169, App
11	142	4.2	943	6	US-10-475-204-34	Sequence 34, Appl
12	142	4.2	1229	7	US-11-054-281-107	Sequence 107, App
13	140.5	4.2	903	6	US-10-689-742-142	Sequence 142, App
14	140.5	4.2	2801	7	US-11-124-368A-305	Sequence 305, App
15	140.5	4.2	2801	7	US-11-124-367A-433	Sequence 433, App
16	140.5	4.2	2896	7	US-11-124-368A-306	Sequence 306, App
17	140.5	4.2	2896	7	US-11-124-367A-434	Sequence 434, App
18	140.5	4.2	3256	7	US-11-124-368A-304	Sequence 304, App
19	140.5	4.2	3256	7	US-11-124-367A-432	Sequence 432, App
20	140	4.2	414	7	US-11-089-551A-28	Sequence 28, Appl
21	138.5	4.1	144	6	US-10-821-234-1254	Sequence 1254, Ap
22	138.5	4.1	1017	7	US-11-054-281-108	Sequence 108, App
23	138.5	4.1	1219	7	US-11-054-281-28	Sequence 28, Appl
24	138.5	4.1	1219	7	US-11-054-281-106	Sequence 106, App
25	138.5	4.1	2804	7	US-11-120-925-3	Sequence 3, Appli
26	138	4.1	697	6	US-10-821-234-905	Sequence 905, App
27	138	4.1	4374	7	US-11-128-572-2	Sequence 2, Appli
28	138	4.1	5024	6	US-10-793-626-2964	Sequence 2964, Ap
29	136	4.1	4384	6	US-10-821-234-1120	Sequence 1120, Ap
30	135.5	4.0	728	6	US-10-530-340-14	Sequence 14, Appl
31	135	4.0	345	7	US-11-024-959-415	Sequence 415, App
32	135	4.0	676	7	US-11-135-855-28	Sequence 28, Appl
33	135	4.0	717	7	US-11-135-855-29	Sequence 29, Appl
34	134.5	4.0	816	7	US-11-072-512-2050	Sequence 2050, Ap
35	134.5	4.0	819	7	US-11-072-512-2755	Sequence 2755, Ap
36	133	4.0	668	6	US-10-453-372-950	Sequence 950, App
37	133	4.0	668	6	US-10-453-372-968	Sequence 968, App
38	133	4.0	668	6	US-10-453-372-970	Sequence 970, App
39	132.5	4.0	8746	7	US-11-098-686-10232	Sequence 10232, A
40	132	3.9	578	6	US-10-821-234-1039	Sequence 1039, Ap
41	132	3.9	1360	7	US-11-241-056-14	Sequence 14, Appl
42	131.5	3.9	251	6	US-10-528-031-8	Sequence 8, Appli
43	131.5	3.9	642	6	US-10-453-372-962	Sequence 962, App
44	131.5	3.9	646	6	US-10-453-372-958	Sequence 958, App
45	131.5	3.9	646	6	US-10-453-372-960	Sequence 960, App

ALIGNMENTS

RESULT 1
US-11-072-512-2651
; Sequence 2651, Application US/11072512
; Publication No. US20060029945A1
; GENERAL INFORMATION:
; APPLICANT: ISOGAI, TAKAO
; APPLICANT: SUGIYAMA, TOMOYASU
; APPLICANT: OTSUKI, TETSUJI
; APPLICANT: WAKAMATSU, AI
; APPLICANT: SATO, HIROYUKI
; APPLICANT: ISHII, SHIZUKO
; APPLICANT: YAMAMOTO, JUN-ICHI
; APPLICANT: ISONO, YUUKO
; APPLICANT: HIO, YURI
; APPLICANT: OTSUKA, KAORU
; APPLICANT: NAGAI, KEIICHI
; APPLICANT: IRIE, RYOTARO
; APPLICANT: TAMECHIKA, ICHIRO
; APPLICANT: SEKI, NAOHIKO
; APPLICANT: YOSHIKAWA, TSUTOMU
; APPLICANT: OTSUKA, MOTOYUKI
; APPLICANT: NAGAHARI, KENJI
; APPLICANT: MASUHO, YASUHIKO
; TITLE OF INVENTION: Novel full length cDNA
; FILE REFERENCE: 084335-0191
; CURRENT APPLICATION NUMBER: US/11/072,512
; CURRENT FILING DATE: 2005-03-07
; PRIOR APPLICATION NUMBER: US 60/350,978
; PRIOR FILING DATE: 2002-01-25
; PRIOR APPLICATION NUMBER: JP 2001-379298
; PRIOR FILING DATE: 2001-11-05
; NUMBER OF SEQ ID NOS: 4096
; SOFTWARE: PatentIn Ver. 2.1
; SEQ ID NO 2651
; LENGTH: 687
; TYPE: PRT
; ORGANISM: Homo sapiens
US-11-072-512-2651

Query Match 4.9%; Score 164.5; DB 7; Length 687;
Best Local Similarity 21.9%; Pred. No. 0.0029;
Matches 74; Conservative 58; Mismatches 109; Indels 97; Gaps 14;

```
Qy      276 KKGMEPLQGHATPALPFKETQELLLSPLPQEGPGSLAAGESSSLSASTSVSDSSQKKEEH 335
            |||  |  | |  | | |:     : :| :|  :   |:::    |    |   :::|
Db       86 KKGATP--GKALVATPGKKG-----AAIPAKGAKN---GKNAKKEDSDEEEDDDSEEDEE 135

Qy      336 NYSLFVSDNLGEQP--------TKCSPEEDEEDEEDVDDEDHDEGFGSEHE--------- 378
```

```
Db        136 DDEDEDEDEDEIEPAAMKAAAAAPASEDEDDEDDEDDEDDDDDEEDDSEEEAMETTPAKG 195

Qy        379 -------------LSENEEEEEEEEDYEDDKDDDISDTFSEPGYENDSVEDLKEVTSISS 425

Db        196 KKAAKVVPVKAKNVAEDEDEEEDDEDEDDDDDEDDEDDDDEDDEEEEEEEEEEPVKEAPG 255

Qy        426 RKRGKRRYFWEYSEQLTPSQQERMLRPSE-------------WNRD-----TLPSNMYQK 467

Db        256 KRKKEMA-----KQKAAPEAKKQKVEGTEPTTAFNLFVGNLNFNKSAPELKTGISDVFAK 310

Qy        468 NGLHHGKYAVKKSRR------TDVEDLTPNPKKLLQI------GNELRKLNKVISDLTPV 515

Db        311 NDLAVVDVRIGMTRKFGYVDFESAEDL----EKALELTGLKVFGNEI-KLEK-------- 357

Qy        516 SELPLTARPRSRKEKNKLAFRACRLKKKAQYEANKVKL 553

Db        358 --------PKGKDSKKERDARTL-LAKNLPYKVTQDEL 386


RESULT 2
US-11-189-817-2
; Sequence 2, Application US/11189817
; Publication No. US20060030006A1
; GENERAL INFORMATION:
;  APPLICANT: INSTITUT PASTEUR
;  APPLICANT:  DRUILHE, PIERRE
;  TITLE OF INVENTION: IDENTIFICATION OF A CONSERVED REGION OF PLASMODIUM FALCIPARUM
;  TITLE OF INVENTION:  MSP3 TARGETED BY BIOLOGICALLY ACTIVE ANTIBODIES
;  FILE REFERENCE: 275601US0
;  CURRENT APPLICATION NUMBER: US/11/189,817
;  CURRENT FILING DATE:  2005-07-27
;  PRIOR APPLICATION NUMBER: 60/598,062
;  PRIOR FILING DATE: 2004-08-03
;  NUMBER OF SEQ ID NOS: 14
;  SOFTWARE: PatentIn version 3.3
; SEQ ID NO 2
;   LENGTH: 354
;   TYPE: PRT
;   ORGANISM: Plasmodium falciparum
US-11-189-817-2

  Query Match             4.9%;  Score 163;  DB 7;  Length 354;
  Best Local Similarity   20.9%;  Pred. No. 0.0016;
  Matches   82;  Conservative   61;  Mismatches  146;  Indels  104;  Gaps   13;

Qy         51 YQQNPRDNFLSLEDCKDIENLESF-TDVLDNEGALTSNWEQWDTYCEDLTKYTKLTSCDI 109

Db         32 YNLNLRNAILN--NNSQIENEENVNTTITGND--FSGGEFLWPGYTEEL-KAKKAS---- 82

Qy        110 WGTKEVDYLGLDDFSSPYQDEEVISKTPTLAQLNSEDSQSVSDSLYYPDSLFSVKQNPLP 169

Db         83 -----------EDAEKAANDAENASK-------EAEEAAKEAVNLKESDKSYT------- 117

Qy        170 SSFPGKKITSRAAAPVCSSKTLQAEVPLSDCVQKASKPPSSTQIMVKTNMYHNEKVNFHV 229

Db        118 ------KAKEACTAASKAKKAVETALKAKDDAEKSSKADS---ISTKTKEYAEKAKNAYE 168

Qy        230 ECKDYVKKAKVKINPVQQSRPLLSQIHTDAAKENTCYCGAVAKRQEKKGMEPLQG-HATP 288

Db        169 KAKNAYQKAN----------------------------QAVLKAKEASSYDYILGWEFGG 200

Qy        289 ALPFKETQELLLSPLPQEGPGSLAAGESSSLSASTSVSDSSQKKEEHNYSLFVSDNLGEQ 348

Db        201 GVPEHKKEENMLSHL-----------------------YVSSKDKENISKENDDVLDEK 236

Qy        349 PTKCSPEEDEEDEEDVDDEDHDEGFGSEHELSENEEEEEEEEDYEDDKDDDISDTFSEPG 408

Db        237 ----EEEAEETEEEELEEKNEEE---TESEISEDEEEEEEEEKEEENDKKKEQEKEQSN 289

Qy        409 YENDSVEDLKEVTSISSRKRGKRRYFWEYSEQL 441

Db        290 ENNDQKKDMEAQNLISKNQNNNEKNVKEAAESI 322


RESULT 3
US-11-128-660-1
; Sequence 1, Application US/11128660
; Publication No. US20060024324A1
; GENERAL INFORMATION:
;  APPLICANT: Statens Serum Institut
;  TITLE OF INVENTION: Vaccines comprising chimeric malaria proteins derived from Plasmoc
;  TITLE OF INVENTION:  falciparum
;  FILE REFERENCE: 15007dk
;  CURRENT APPLICATION NUMBER: US/11/128,660
;  CURRENT FILING DATE:  2005-05-12
;  NUMBER OF SEQ ID NOS: 2
;  SOFTWARE: PatentIn version 3.1
; SEQ ID NO 1
;   LENGTH: 651
;   TYPE: PRT
;   ORGANISM: Plasmodium falciparum
US-11-128-660-1

  Query Match             4.7%;  Score 157.5;  DB 7;  Length 651;
  Best Local Similarity   18.4%;  Pred. No. 0.0076;
  Matches   99;  Conservative   99;  Mismatches  171;  Indels  169;  Gaps   23;

Qy         25 QTLMSTDLLANSSDPDFMYELDRE-MNYQQNPRDNFLSLEDCKDIENLESFTDVL----- 78

Db        130 ESLSDLELLENSSQDN----LDKDTISTEPFPNQKHKDLQQDLNDEPLEPFPTQIHKDYK 185

Qy         79 --------DNEGALTSNWEQWDTYCEDLTKYTKLTSCDIWGTKEVDYLGLDDFSSPYQDE 130

Db        186 EKNLINEEDSEPFPRQEHKKVDNHNEEKNVFHENGSAN--GNQ--GSLKLKSFDEHLKDE 241

Qy        131 EV-----------ISKTPTLAQLNS-EDSQSVSDSLYYPDSLFSVKQNPLPSSFPGKKIT 178

Db        242 KIENEPLVHENLSIPNDPIEQILNQPEQETNIQEQLYNEKQNVEEKQN------------ 289

Qy        179 SRAAAPVCSSKTLQAEVPLSDCVQKASKPPSSTQIMVKTNMYHNEK-----VNFHVE--- 230

Db        290 --------------SQIPSLDL-----KEPTNEDILPNHNPLENIKQSESEIN-HVQDHA 329

Qy        231 ------------CKDYVKKAKVKINPVQ-----------QSRP----------------- 250

Db        330 LPKENIIDKLDNQKEHIDQSQHNINVLQENNINNHQLEPQEKPNIESFEPKNIDSEIILP 389

Qy        251 -------LLSQIHTDAAKENTCYCGAVAKRQEKKGMEPLQGHATPALPFKETQELLLSPL 303

Db        390 ENVETEEIIDDVPSPKHSNHETFEEETSESEHEEAVSEKNAHET--VEHEETVSQESNPE 447

Qy        304 PQEGPGSLAAGESSSLSASTSV-SDSSQ---------------------------KKEE 334

Db        448 KADNDGNVSQNSNNELNENEFVESEKSEHEARSKAKEASSYDYILGWEFGGGVPEHKKEE 507

Qy        335 HNYS-LFVS----DNLGEQPTKCSPEEDEE----DEEDVDDEDHDEGFGSEHELSENEEE 385

Db        508 NMLSHLYVSSKDKENISKENDDVLDEKEEEAEETEEEELEEKNEEE---TESEISEDEEE 564

Qy        386 EEEEEDYEDD--KDDDISDTFSEPGYENDSVEDLKEVTSISSRKRGKRRYFWEYSEQL 441

Db        565 EEEEEKEEENEKKKEQEKEQSNE---NNDQKKDMEAQNLISKNQNNNEKNVKEAAESI 619


RESULT 4
US-10-509-422-2
; Sequence 2, Application US/10509422
; Publication No. US20050244825A1
; GENERAL INFORMATION:
;  APPLICANT: Liou, Simon
;  TITLE OF INVENTION: Human BMP2 Inducible Kinases
;  FILE REFERENCE: 004974.01015
;  CURRENT APPLICATION NUMBER: US/10/509,422
;  CURRENT FILING DATE:  2004-09-24
;  PRIOR APPLICATION NUMBER: PCT/EP03/080825
```

```
;  PRIOR FILING DATE: 2003-03-20
;  PRIOR APPLICATION NUMBER: US 60/367,512
;  PRIOR FILING DATE: 2002-03-27
;  PRIOR APPLICATION NUMBER: US 60/406,936
;  PRIOR FILING DATE: 2002-08-30
;  NUMBER OF SEQ ID NOS: 9
;  SOFTWARE: FastSEQ for Windows Version 4.0
; SEQ ID NO 2
;   LENGTH: 1168
;   TYPE: PRT
;   ORGANISM: Homo sapiens
US-10-509-422-2

  Query Match             4.6%;  Score 155;  DB 6;  Length 1168;
  Best Local Similarity   18.8%;  Pred. No. 0.023;
  Matches  117;  Conservative   81;  Mismatches  189;  Indels  236;  Gaps   26;

Qy          2 PQPSVS---GMDPPFGDAFRSHTFSEQTLMSTDLLANSSDPDFMYELDR---------- 47
              |||| |    | | :  ||    | :| |         : |:::
Db        526 PQPSASQYPTMMPQYQQAF----FQQQMLAQHQPSQQQASPEYLTSPQEFSPALVSYTSS 581

Qy         48 ---------EMNYQQNPRDNFLSLEDCKDIENL---ESFTDVLDNEGALTSNWEQWDTYC 95
                      : :|  | :  | |: | : | |    :: ::  | |        |: :
Db        582 LPAQVGTIMDSSYSANRQVFFQSVADKEAIANFTNQKNISNPPDMSG--------WNPFG 633

Qy         96 EDLTKYTKLTSCDIWGTKEVDYLGLDDFSSPYQDEEVISKTPTLAQLNSEDSQSVSDSLY 155
              ||   ::|||                :||:: :   | : | : :: ||
Db        634 ED--NFSKLT----------------------EEELLDREFDLLRSNRLEERASSDK-- 666

Qy        156 YPDSLFSVKQNPLPSSFPGKKITSRAAAPVCSSKTLQAEVPLSDCVQKASKPPSSTQIMV 215
              |||  :   :|    |                      ||
Db        667 NVDSLSAPHNHPPEDPF-------------------GSVPF------------------ 688

Qy        216 KTNMYHNEKVNFHVECKDYVKKAKVKINPVQQSRPLLSQIHTDAAKENTCYCGAVAKRQE 275
                : |: |                                              |:  |: |
Db        689 ---ISHSGK-------------------------------------------GSPEKKAE 702

Qy        276 KKGMEPLQGHATPALPFKETQELLLSPLPQEGPGSLAAGESSSLSASTSVSDSSQKKEEH 335
                :    | | |    |       ||         | :         |||    ||  :
Db        703 HSSINQENGTANPIKNGK------TSP----------ASKDQRTGKKTSVQGQVQKGNDE 746

Qy        336 NYSLFVSDNLGEQPTKCSPEEDEEDEEDV---------DDEDHDEGFGSEHELSENEEEE 386
              : | | ||      |:  | ||:|:|:|:|         ||:   |  |    | ::|:||
Db        747 SESDFESD----PPSPKSSEEEEQDDEEVLQGEQGDFNDDDTEPENLGHRPLLMDSEDEE 802

Qy        387 EEEE-----DYEDDKD--DDISDTFSEPGYENDSVEDLKEVTSISSRKRGKRRYFWEYSE 439
              |||:     |||  |    |:|  : :    :   :|| :   |:          : |
Db        803 EEEKHSSDSDYEQAKAKYSDMSSVYRDRS-GSGPTQDLNTILLTSA----------QLSS 851

Qy        440 QL---TPSQQ-----------ERMLRP-SEWNRDTLPSNMYQKNGLHHGKYAV----KKS 480
              :   || |:           |  :|  | |   || : :   ||   ::  |      |
Db        852 DVAVETPKQEFDVFGAVPFFAVRAQQPQQEKNEKNLPQHRFPAAGLEQEEFDVFTKAPFS 911

Qy        481 RRTDVED----------LTPNPKKLLQIGN----------ELRKLNKVISDLTPVSELPL 520
              :: :|::          :   || :   |:             : |: :  | |  |:
Db        912 KKVNVQECHAVGPEAHTIPGYPKSVDVFGSTPFQPFLTSTSKSESNEDLFGLVPFDEITG 971

Qy        521 TARPR-SRKEKNKLAFRACRLKK 542
              : : :  ::   ||: |  | |:
Db        972 SQQQKVKQRSLQKLSSRQRRTKQ 994


RESULT 5
US-11-124-367A-293
; Sequence 293, Application US/11124367A
; Publication No. US20060024700A1
; GENERAL INFORMATION:
;  APPLICANT: Michele Cargill
;  APPLICANT:  Hongjin Huang
;  TITLE OF INVENTION: Genetic Polymorphisms Associated with
;  TITLE OF INVENTION:  Fibrosis Methods of Detection and Uses Thereof
;  FILE REFERENCE: CL001519.ORD
;  CURRENT APPLICATION NUMBER: US/11/124,367A
;  CURRENT FILING DATE:  2005-05-09
;  PRIOR APPLICATION NUMBER: US 60/568,846
;  PRIOR FILING DATE: 2004-05-07
;  PRIOR APPLICATION NUMBER: US 60/582,609
;  PRIOR FILING DATE: 2004-06-25
;  PRIOR APPLICATION NUMBER: US 60/599,554
;  PRIOR FILING DATE: 2004-08-09
;  NUMBER OF SEQ ID NOS: 34460
;  SOFTWARE: FastSEQ for Windows Version 4.0
; SEQ ID NO 293
;   LENGTH: 740
;   TYPE: PRT
;   ORGANISM: Homo sapiens
US-11-124-367A-293

  Query Match             4.6%;  Score 154;  DB 7;  Length 740;
  Best Local Similarity   22.0%;  Pred. No. 0.015;
  Matches  100;  Conservative   63;  Mismatches  162;  Indels  130;  Gaps   20;

Qy        113 KEVDYLGLDDFSSPYQDE-----EVISKTPTLAQLNSEDSQSVSDSLY----------YP 157
              ||:|   |||  | |  |      ::|      | :|  :|  |::  :           ||
Db        208 KELDLSELDDPDSAYLQEARLKRKLIRLFGRLCEL--KDCSSLTGRVIEQRIPYRGTRYP 265

Qy        158 DSLFSVKQ---NPLPSSFPGKKITSRAAAPVCSSKTL-----QAEVPLSDCVQKASKPPS 209
              :    :::    | | :||       ||      : :|     | ::   |  :
Db        266 EVNRRIERLINKPGPDTFPDYGDVLRAVEKAAARHSLGLPRQQLQLMAQDAFRDVG---- 321

Qy        210 STQIMVKTNMYHNEKVNFHVECKDYVKKAKVKINPVQQSRPLLSQIHTDAAKEN------ 263
                | ::   : :   ||      |  :     ::|   | |:|::      :||
Db        322 ---IRLQERRHLDLIYNFGCHLTDDYRPG---VDPA-LSDPVLAR----RLRENRSLAMS 370

Qy        264 ------TCYCGAVAKRQE---KKGMEPLQGHATPALPFKETQELLLSPLPQEGPGSLAA- 313
                   : |      | :|   ||     ||| :: :    :| |  |     |||  :|:
Db        371 RLDEVISKYAMLQDKSEEGERKKRRARLQGTSSHS---ADTPEASLD--SGEGPSGMASQ 425

Qy        314 GESSSLSASTSVSDSSQKKEEHNYSLFVSDNLGEQPTKCSPEEDEEDEEDVDDEDHDEGF 373
              |  |:  | |   |  :  ||                    ||:||:||:  | : :|
Db        426 GCPSASRAETDDEDDEESDEEEE-----------------EEEEEEEEEATDSEEEEDL 467

Qy        374 GSEHELSENEEEEEEEEDYEDDKDDDISDTFSEPGYENDSVEDLKEVTSISSRKRGKRRY 433
              |  |:::|||:|||:     || | |   |        ::|  |:::  |  :: | |
Db        468 EQMQEGQEDDEEEDEEEEAAAGKDGDKSPMSSLQISNEKNLEPGKQISRSSGEQQNKGRI 527

Qy        434 FWEYSEQLTPS-QQERMLRPSEWNRDTLPSNMYQKNGLHHGKYAVKKSRRTDVEDLTPNP 492
                  ::||    |  | ||                                | |    |
Db        528 -------VSPSLLSEEPLAPSS---------------------------IDAESNGEQP 552

Qy        493 KKLLQIGNELRKLNKVISDLTPVS---ELPLTARP 524
              ::|             : : :|||   || : | |
Db        553 EEL-----------TLEEESPVSQLFELEIEALP 575


RESULT 6
US-11-124-367A-292
; Sequence 292, Application US/11124367A
; Publication No. US20060024700A1
; GENERAL INFORMATION:
;  APPLICANT: Michele Cargill
;  APPLICANT:  Hongjin Huang
;  TITLE OF INVENTION: Genetic Polymorphisms Associated with
;  TITLE OF INVENTION:  Fibrosis Methods of Detection and Uses Thereof
;  FILE REFERENCE: CL001519.ORD
;  CURRENT APPLICATION NUMBER: US/11/124,367A
;  CURRENT FILING DATE:  2005-05-09
;  PRIOR APPLICATION NUMBER: US 60/568,846
;  PRIOR FILING DATE: 2004-05-07
;  PRIOR APPLICATION NUMBER: US 60/582,609
;  PRIOR FILING DATE: 2004-06-25
;  PRIOR APPLICATION NUMBER: US 60/599,554
;  PRIOR FILING DATE: 2004-08-09
;  NUMBER OF SEQ ID NOS: 34460
;  SOFTWARE: FastSEQ for Windows Version 4.0
; SEQ ID NO 292
```

```
;   LENGTH: 760
;   TYPE: PRT
;   ORGANISM: Homo sapiens
US-11-124-367A-292

  Query Match             4.6%;  Score 154;  DB 7;  Length 760;
  Best Local Similarity   22.0%;  Pred. No. 0.015;
  Matches  100;  Conservative   63;  Mismatches  162;  Indels  130;  Gaps   20;

Qy        113 KEVDYLGLDDFSSPYQDE-----EVISKTPTLAQLNSEDSQSVSDSLY----------YP 157
              ||:|   |||  | |  |     ::|     | :|  :|  |::  :            ||
Db        228 KELDLSELDDPDSAYLQEARLKRKLIRLFGRLCEL--KDCSSLTGRVIEQRIPYRGTRYP 285

Qy        158 DSLFSVKQ---NPLPSSFPGKKITSRAAAPVCSSKTL-----QAEVPLSDCVQKASKPPS 209
              :   :::    | | :||      ||     :  :|     | ::    |  :
Db        286 EVNRRIERLINKPGPDTFPDYGDVLRAVEKAAARHSLGLPRQQLQLMAQDAFRDVG---- 341

Qy        210 STQIMVKTNMYHNEKVNFHVECKDYVKKAKVKINPVQQSRPLLSQIHTDAAKEN------ 263
                | ::   : :   ||     |  :     ::|   | |:|::      :||
Db        342 ---IRLQERRHLDLIYNFGCHLTDDYRPG---VDPA-LSDPVLAR----RLRENRSLAMS 390

Qy        264 ------TCYCGAVAKRQE---KKGMEPLQGHATPALPFKETQELLLSPLPQEGPGSLAA- 313
                   : |     | :|   ||    ||| :: :    :| | |     |||  :|:
Db        391 RLDEVISKYAMLQDKSEEGERKKRRARLQGTSSHS---ADTPEASLD--SGEGPSGMASQ 445

Qy        314 GESSSLSASTSVSDSSQKKEEHNYSLFVSDNLGEQPTKCSPEEDEEDEEDVDDEDHDEGF 373
              |  |:  | |   |  :  ||                  ||:||:||:  | : :|
Db        446 GCPSASRAETDDEDDEESDEEEE-----------------EEEEEEEEATDSEEEEDL 487

Qy        374 GSEHELSENEEEEEEEEDYEDDKDDDISDTFSEPGYENDSVEDLKEVTSISSRKRGKRRY 433
                |  |::|||:|||:    || | |   |        ::|  |:::  |  :: | |
Db        488 EQMQEGQEDDEEEDEEEEAAAGKDGDKSPMSSLQISNEKNLEPGKQISRSSGEQQNKGRI 547

Qy        434 FWEYSEQLTPS-QQERMLRPSEWNRDTLPSNMYQKNGLHHGKYAVKKSRRTDVEDLTPNP 492
                  ::||   |  | ||                                | |     |
Db        548 -------VSPSLLSEEPLAPSS--------------------------IDAESNGEQP 572

Qy        493 KKLLQIGNELRKLNKVISDLTPVS---ELPLTARP 524
              ::|             : : :|||   || : | |
Db        573 EEL-----------TLEEESPVSQLFELEIEALP 595


RESULT 7
US-10-984-645-2
; Sequence 2, Application US/10984645
; Publication No. US20050244386A1
; GENERAL INFORMATION:
;  APPLICANT: Habener, Joel
;  APPLICANT:  Zulewski, Hendrik
;  APPLICANT:  Abraham, Elizabeth
;  APPLICANT:  Vallejo, Mario
;  TITLE OF INVENTION: METHOD OF TRANSPLANTING IN A MAMMAL AND TREATING DIABETES MELLITU
;  TITLE OF INVENTION:  BY ADMINISTERING A PSEUDO-ISLET LIKE AGGREGATE DIFFERENTIATED FR
;  TITLE OF INVENTION:  A NESTIN-POSITIVE PANCREATIC STEM CELL
;  FILE REFERENCE: 3284/1223
;  CURRENT APPLICATION NUMBER: US/10/984,645
;  CURRENT FILING DATE:  2004-11-09
;  PRIOR APPLICATION NUMBER: US 09/731,255
;  PRIOR FILING DATE: 2000-12-06
;  PRIOR APPLICATION NUMBER: US 60/169,082
;  PRIOR FILING DATE: 1999-12-06
;  PRIOR APPLICATION NUMBER: US 60/215,109
;  PRIOR FILING DATE: 2000-06-28
;  PRIOR APPLICATION NUMBER: US 60/239,880
;  PRIOR FILING DATE: 2000-10-06
;  NUMBER OF SEQ ID NOS: 55
;  SOFTWARE: PatentIn version 3.2
; SEQ ID NO 2
;   LENGTH: 1618
;   TYPE: PRT
;   ORGANISM: Homo sapiens
US-10-984-645-2

  Query Match             4.4%;  Score 147.5;  DB 6;  Length 1618;
  Best Local Similarity   19.9%;  Pred. No. 0.1;
  Matches  116;  Conservative   90;  Mismatches  204;  Indels  173;  Gaps   24;

Qy        124 SSPYQDEEV---ISKTPTLAQLNSEDSQSVSDSLYYPDSLFSVKQNPLPSSFPGKKITSR 180
              |  :|| ::     :||   :| |    |     |  | :  : |:|:    ::   :
Db        322 SLSFQDPKLELQFPRTPEGRRLGS--LLPVLSPTSLPSPLPATLETPVPAFLKNQEFL-Q 378

Qy        181 AAAPVCSS----KTLQAEVPLSDCVQKASKPPSSTQIMVKTNMYHNEKVNFHVECKDYVK 236
              |  | :|     | ||  |  |   :|  |
Db        379 ARTPTLASTPIPPTPQAPSPAVDAEIRAQDAP--------------------------- 410

Qy        237 KAKVKINPVQQSRPLLSQIHTDAAKENTCYCGAVAKRQEKKGMEPLQGHATPALPFK--- 293
                              || : |   :               |:  |||:  |  | |:|
Db        411 --------------LSLLQTQGGR-------------KQAPEPLRAEARVAIPASVLP 441

Qy        294 -------ETQELLLSPLPQEGPGSLA---AGESSSLSASTSVSDSSQ-------KKEEHN 336
                    : ||      | |   |||   : : ||| |     |  |:       : |
Db        442 GPEEPGGQRQEASTGQSP-EDHASLAPPLSPDHSSLEAKDGESGGSRVFSICRGEGEGQI 500

Qy        337 YSLFVSDNLGEQPTKCSPEEDEEDEEDVD-DEDHDEGFGSEHEL--SENEEEEEEEEDYE 393
              : |   :   |     | :::  :|||::  |  |     | |   |  || :|  :  |
Db        501 WGLVEKETAIEGKVVSSLQQEIWEEEDLNRKEIQDSQVPLEKETLKSLGEEIQESLKTLE 560

Qy        394 DDKDDDISDTFSE-PGYENDSVEDLKEVT-----------SISSRKRGKRRYFWEYSEQ 440
              :   : :     | |    : :| || :              | :|||| |         |
Db        561 NQSHETLERENQECPRSLEEDLETLKSLEKENKRAIKGCGGSETSRKRGCR--------Q 612

Qy        441 LTPS------------QQERMLRPSEWNRDT--LPSNMYQKNGLHHGKYAVKKSRRTDV 485
              | |:              : : :::  | | :|   |     |:              :  | : ::
Db        613 LKPTGKEDTQTLQSLQKENQELMKSLEGNLETFLFPGTENQE---------LVSSLQENL 663

Qy        486 EDLT------PNPKKLLQIGNELRKLNKVISDLTPVSELPLTARPRSRKEKNKLAFRACR 539
              | ||        | :  ::|:|     : :  ||  :: ||    || :::|| |||:
Db        664 ESLTALEKENQEPLRSPEVGDE-----EALRPLTKENQEPL----RSLEDENKEAFRS-- 712

Qy        540 LKKKAQ------YEANKVKLWGLNTEYDNLLFVINSIKQEIVNRVQNPRDERGPNMGQKL 593
              |:|: |        | ::  :  | ||     |  :    || :  ::    :|  ::|::
Db        713 LEKENQEPLKTLEEEDQSIVRPLETENHKSLRSLEEQDQETLRTLEKETQQRRRSLGEQD 772

Qy        594 EILIK--------------DTLGLPVAGQTSEFVNQVLEKTAE 622
              :: ::              :  |:  :  ||:  : |:: |
Db        773 QMTLRPPEKVDLEPLKSLDQEIARPLENENQEFLKSLKEESVE 815


RESULT 8
US-10-770-726-49
; Sequence 49, Application US/10770726
; Publication No. US20050266409A1
; GENERAL INFORMATION:
;  APPLICANT: Wyeth
;  APPLICANT:  Brown, Eugene
;  APPLICANT:  Liu, Wei
;  TITLE OF INVENTION: COMPOSITIONS AND METHODS FOR DIAGNOSING, PREVENTING, AND TREATING
;  TITLE OF INVENTION:  CANCERS
;  FILE REFERENCE: AM101079 (031896-010000)
;  CURRENT APPLICATION NUMBER: US/10/770,726
;  CURRENT FILING DATE:  2004-02-04
;  NUMBER OF SEQ ID NOS: 48640
;  SOFTWARE: PatentIn version 3.2
; SEQ ID NO 49
;   LENGTH: 795
;   TYPE: PRT
;   ORGANISM: Homo sapiens
US-10-770-726-49

  Query Match             4.3%;  Score 145.5;  DB 6;  Length 795;
  Best Local Similarity   20.2%;  Pred. No. 0.056;
  Matches  116;  Conservative   80;  Mismatches  214;  Indels  165;  Gaps   25;

Qy        108 DIWGTKEVDYLGLDDFSSPYQDEEVISKTPTLAQLNSEDSQSVSDSLYYPDSLFSVKQNP 167
              | |  | :| : | :     : ||   |          ||:|  |  |||          ::
Db          6 DSWKVKTLDEI-LQEKKRRKEQEE---KAEIKRLKNSDDRDSKRDSL---------EEGE 52
```

```
Qy        168 LPSSFPGKKITSRAAAPVCSSKTLQAEVPLSDCVQKASKPPSSTQIMVKTNMYH------ 221
Db         53 LRDH--RMEITIR-NSPYRREDSMEDRGEEDDSL--AIKPPQ--QMSRKEKVHHRKDEKR 105

Qy        222 ------------------------NEKVNFHVECKDYV-------KKAKVKINPVQQSR 249
Db        106 KEKRRHRSHSAEGGKHARVKEKEREHERRKRHREEQDKARREWERQKRREMAREHSRRER 165

Qy        250 PLLSQIHTDAAKENTC-------------YCGAVAKRQEKKGMEPLQGH---------- 285
Db        166 DRLEQLERKRERERKMREQQKEQREQKERERRAEERRKEREARREVSAHHRTMREDYSDK 225

Qy        286 -------ATPALPFKETQELLLSPLPQEGPGSLAAGESSSLSASTSVSDSSQKKEEHNYS 338
Db        226 VKASHWSRSPPRPPRERFEL--------GDGR-KPGEARPARAQKPAQLKEEKMEERD-- 274

Qy        339 LFVSD--NLGEQPTKCSPEED--------EEDEEDVDDEDHDEGFGSEHELSENEEEEEE 388
Db        275 -LLSDLQDISDSERKTSSAESSSAESGSGSEEEEEEEEEEEEEGSTSE----ESEEEEEE 329

Qy        389 EEDYEDDKDDDISDTFSEPGYENDSVEDLKEVTSISSRKRGKRRYFWEYSEQLTPSQQER 448
Db        330 EEEEEEETGSN-SEEASEQSAEEVSEEEMSE-----DEERENENHLLVVPE----SRFDR 379

Qy        449 MLRPSEWNRDTLPSNMYQKNGLHHGKYAVKKSRRTDVEDLTPNPKKLLQIGNELRKLNKV 508
Db        380 DSGESEEAEEEVGEGTPQSSALTEGDYV------PDSPALSP-----IELKQELPKYLPA 428

Qy        509 ISDLTPVSEL----------------------PLTARPRSRKEKNKLAFRACRLKK--- 542
Db        429 LQGCRSVEEFQCLNRIEEGTYGVVYRAKDKKTDEIVALKRLKMEKEKEGFPITSLREINT 488

Qy        543 --KAQY-EANKVKLWGLNTEYDNLLFVINSIKQEI 574
Db        489 ILKAQHPNIVTVREIVVGSNMDKIYIVMNYVEHDL 523
```

RESULT 9
US-10-495-083-3
; Sequence 3, Application US/10495083
; Publication No. US20060008800A1
; GENERAL INFORMATION:
; APPLICANT: Max-Planck-Gesellschaft zur Forderung der Wissenschaften e.V.
; APPLICANT: ROSENMUND, Christian
; APPLICANT: BROSE, Nils
; APPLICANT: RHEE, Jeong-Seop
; APPLICANT: BETZ, Andrea
; APPLICANT: RETTIG, Jens
; APPLICANT: ASHERY, Uri
; APPLICANT: JUNGE, Harald
; TITLE OF INVENTION: UNC-13 IN THE MODULATION OF NEUROTRANSMISSION AND SECRETION EVENT
; FILE REFERENCE: 009848-0309442
; CURRENT APPLICATION NUMBER: US/10/495,083
; CURRENT FILING DATE: 2004-05-05
; PRIOR APPLICATION NUMBER: PCT/EP02/12072
; PRIOR FILING DATE: 2002-10-29
; PRIOR APPLICATION NUMBER: EP 01 12 6235.9
; PRIOR FILING DATE: 2001-11-05
; NUMBER OF SEQ ID NOS: 12
; SOFTWARE: PatentIn version 3.1
; SEQ ID NO 3
; LENGTH: 1735
; TYPE: PRT
; ORGANISM: Rattusnorvegicus
US-10-495-083-3

Query Match 4.3%; Score 145.5; DB 6; Length 1735;
Best Local Similarity 20.4%; Pred. No. 0.15;
Matches 142; Conservative 87; Mismatches 208; Indels 259; Gaps 37;

```
Qy         21 TFSEQTLMSTDLLANSSDP----DFMYELDREMNYQQNPRDNFLSLEDCKDIENLESFTD 76
Db         25 TLKVQNVKSTTIAVRGSQPSWEQDFMFEINR--------LDLGLTVE------------- 63

Qy         77 VLDNEGALTSNWEQWDTYC------------------EDLTKYTK--LTSCDIWGTKEV 115
Db         64 -VWNKGLI------WDTMVGTVWIPLRTIRQSNEEGPGEWLTLDSQAIMADSEICGTKDP 116

Qy        116 DY--LGLD-DFSSPYQ--DEEVISKTPTLAQLNS---EDSQSVSDSLYYPDSLFSVKQNP 167
Db        117 TFHRILLDAHFELPLDIPEEEARYWAKKLEQLNAMRDQDEYSFQDQ----------QDKP 166

Qy        168 LPSSFPGKKITSRAAAPVCSSKTLQAEVPLSDCVQ-------KASKPPSSTQIMVKTNMY 220
Db        167 LP------------------------VPSSQCCNWNYFGWGEQNDDPDSA-VDDRDSDY 200

Qy        221 HNEKVN-----FHVECKDYVKKAKVKINPVQQSRPLLSQIHTDAAKENTCYCGAVAKRQE 275
Db        201 RSETSNSIPPPYYTTSQPNASVHQYSVRP----PPLGSR---------ESYSDSMHSYEE 247

Qy        276 KKGMEPLQGHATPALPFKETQELLLSPLPQEGPGSLAAGESSSLSASTSVSDSSQKKEEH 335
Db        248 ----------------FSEPR--ALSP-----TGSSRYASSGELSQGSSQLSEDFDPDEH 284

Qy        336 NYSLFVSDNLGEQPT-----------KCSP--EEDEEDEEDVDDEDHDEGFGSEHELSE- 381
Db        285 SLQGSELDDERDRDSYHSCHSSVSYHKDSPRWDQDEEDLEDLEDLE-DEELPEEEELEEE 343

Qy        382 --NEEEEEEEEDYEDDKDDDISDTFSEPGYENDSVEDLKEVTSISSRKRGKRRYFWEYSE 439
Db        344 ELEEEEELEEEELELEEEEEVPDDLA----------------------------SYTQ 373

Qy        440 QLTPSQQERMLRPSEWNRDTLPSNMYQKNGLHHGKYAVKKSRRTDV-----EDLTPNPKK 494
Db        374 Q----EDTTVAEPKEFKRISFPTAAPQK----EDKVSAVPIEAPDVSKGIPKAATPEEKA 425

Qy        495 LLQIGNELRKLNKVISDLTPVSELPLTAR----------PRSRKEKNKL-AFRACRLK-K 542
Db        426 AAECAQEAE---------PPKSEESFRSREAEEGQEGQDAMSRAKANWLRAFNKVRMQLQ 476

Qy        543 KAQYEANKVK-LWGLNTEYDNLLFVINSIKQEIVNRVQNPRDERGPNMGQKLEILIKDTL 601
Db        477 EARGEGEMSKSLWFKGGPGGGLI-IIDSM----------------PDIRKRKPIPLVSDL 519

Qy        602 GLPV-----AGQTS----------EFVNQVLEKTAE 622
Db        520 AMSLVQSRKAGITSALASSTLNNEELKNHVYKKTLQ 555
```

RESULT 10
US-10-878-556A-169
; Sequence 169, Application US/10878556A
; Publication No. US20050266399A1
; GENERAL INFORMATION:
; APPLICANT: Hoffmann La-Roche Inc.
; TITLE OF INVENTION: HCV regulated protein expression
; FILE REFERENCE: 21762
; CURRENT APPLICATION NUMBER: US/10/878,556A
; CURRENT FILING DATE: 2004-06-28
; NUMBER OF SEQ ID NOS: 199
; SOFTWARE: PatentIn version 3.1
; SEQ ID NO 169
; LENGTH: 1404
; TYPE: PRT
; ORGANISM: Homo sapiens
; PUBLICATION INFORMATION:
; DATABASE ACCESSION NUMBER: humangp/chr12-q14221
; DATABASE ENTRY DATE: 2003-04-22
US-10-878-556A-169

Query Match 4.3%; Score 144; DB 6; Length 1404;
Best Local Similarity 19.8%; Pred. No. 0.14;
Matches 141; Conservative 132; Mismatches 268; Indels 172; Gaps 32;

```
Qy         18 RSHTFSEQTLMSTDLLANSSD------PDFMYELDREMNYQQNPRDNFLSLEDCKDIENL 71
```

```
Db        562 KNHTLQEQVTQLTEKLKNQSESHKQAQENLHDQVQEQKAHLRAAQDRVLSLE-------- 613

Qy         72 ESFTDVLDNEGALTSNWE---QWDTYCEDLTKYTKLTSCDIWGTKEVDYL--GLDDFSSP 126

Db        614 ---TSVNELNSQLNESKEKVSQLDIQIKAKTEL--LLSAEAAKTAQRADLQNHLDTAQNA 668

Qy        127 YQD--EEVISKTPTLAQLNS--EDSQSVSDSL-----YYPDSLFSVKQNPLPSSFPGKKI 177

Db        669 LQDKQQELNKITTQLDQVTAKLQDKQEHCSQLESHLKEYKEKYLSLEQKTEELEGQIKKL 728

Qy        178 TSRA----AAPVCSSKTLQAEVPL-SDCVQKASKPPSSTQIMVKTNMYHNEKVNFHVECK 232

Db        729 EADSLEVKASKEQALQDLQQQRQLNTDLELRATE--LSKQLEMEKEIVSSTRLDLQKK-S 785

Qy        233 DYVKKAKVKINPVQQSRPLLSQIHTDAAKENTCYCGAVAKRQEKKGMEPLQGHATPALPF 292

Db        786 EALESIKQKLTKQEEEKKILKQDFETLSQE--------TKIQHEELNNRIQTTVTELQKV 837

Qy        293 KETQELLLSPLPQEGPGSLAAGESSSLSASTSVSDSSQKKE----------------E 334

Db        838 KMEKEALMTEL------STVKDKLSKVSDSLKNSKSEFEKENQKGKAAILDLEKTCKELK 891

Qy        335 HNYSLFVSDNLGEQ-PTKCSPEEDEEDEEDVD---DEDHDEGFGSEHELSENEEEEEEEE 390

Db        892 HQLQVQMENTLKEQKELKKSLEKEKEASHQLKLELNSMQEQLIQAQNTLKQNEKEEQQLQ 951

Qy        391 DYEDDKDDDISDTFSEPGYENDSVEDLKEVTSISSRKRGKRRYFWEYSEQLTPSQQERML 450

Db        952 -------GNINELKQSSEQKKKQIEALQGELKIAVLQKTELEN--KLQQQLTQAAQE--- 999

Qy        451 RPSEWNRDTLPSNMYQKN--------GLHHGKYAVKKSRRTDV----------EDLTPN 491

Db       1000 LAAEKEKISVLQNNYEKSQETFKQLQSDFYGRESELLATRQDLKSVEEKLSLAQEDLISN 1059

Qy        492 PKKLLQIGNELRKLNKVISDLTPVSELPLTARPRSRKEKNKLAFRACR----------LK 541

Db       1060 RN---QIGNQ----NKLIQELKTAK---ATLEQDSAKKEQQLQER-CKALQDIQKEKSLK 1108

Qy        542 KK---------AQYEANKVK----LWGLNTEYDNLLFVINSIKQEIVNRVQNPRDERGPN 588

Db       1109 EKELVNEKSKLAEIEEIKCRQEKEITKLNEE-------LKSHKLESIKEITNLKDAKQLL 1161

Qy        589 MGQKLEI------------------LIKDTLGLPVAGQTSEFVNQVLEKTAE 622

Db       1162 IQQKLELQGKADSLKAAVEQEKRNQQILKD----QVKKEEEELKKEFIEKEAK 1210


RESULT 11
US-10-475-204-34
; Sequence 34, Application US/10475204
; Publication No. US20050277116A1
; GENERAL INFORMATION:
;  APPLICANT: PRESIDENT AND FELLOWS OF HARVARD COLLEGE
;  TITLE OF INVENTION: COMPOSITIONS AND METHODS FOR THE IDENTIFICATION OF
;  TITLE OF INVENTION:  PROTEIN INTERACTIONS IN VERTEBRATE CELLS
;  FILE REFERENCE: HMV-056.25
;  CURRENT APPLICATION NUMBER: US/10/475,204
;  CURRENT FILING DATE:  2003-10-17
;  PRIOR APPLICATION NUMBER: PCT/US02/13008
;  PRIOR FILING DATE: 2002-09-27
;  PRIOR APPLICATION NUMBER: 60/285,509
;  PRIOR FILING DATE: 2001-04-20
;  NUMBER OF SEQ ID NOS: 35
;  SOFTWARE: PatentIn Ver. 2.1
; SEQ ID NO 34
;   LENGTH: 943
;   TYPE: PRT
;   ORGANISM: Homo sapiens
US-10-475-204-34

  Query Match             4.2%;  Score 142;  DB 6;  Length 943;
  Best Local Similarity   22.2%;  Pred. No. 0.12;
  Matches  110;  Conservative   71;  Mismatches  147;  Indels  168;  Gaps   30;

Qy        131 EVISK-TP---TLAQLNSEDSQSVSDSLYY---------PDSLFSVKQNPLPSSFPGKKI 177

Db        124 DVSSKNTPDSKKISSRNINDHHSEADEEFYLSVGSPSVLLDAKTSVSQNVIPSS------ 177

Qy        178 TSRAAAPVCSSKTLQAEVPLSDCVQKASKPPSSTQIMVKTNMYHNEKVNF-------HVE 230

Db        178 -----AKKRETYTFENSVNML---------PSSTEVSVKT----KKRLNFDDKVMLKKIE 219

Qy        231 CKDYVKKAKVKINPVQQSRPLLS----------QIHTDAAKE-NTCYCGAVAKRQEKKGM 279

Db        220 IDNKVSDEEDKTSEGQERKPSGSSQNRIRDSEYEIQRQAKKSFSTLFLETVKRKSES--- 276

Qy        280 EPLQGHATPALP---------------FKETQELLLS----PLPQEGPGSLAAGESSSL 319

Db        277 SPIVRHAATAPPHSCPPDDTKLIEDEFIIDESDQSFASRSWITIPRKA-GSLKQ-RTISP 334

Qy        320 SASTSVSDSSQKKEEHNYSL---FVSDNLGEQPTKCSPEEDEE--DEEDVD------DED 368

Db        335 AESTALFQGRKSREKHHNILPKTLANDKHSHKP---HPVETSQPSDKTVLDTSYALIDET 391

Qy        369 HDEGFGSEHEL-SENEEE--------------------EEEEEDYEDDKDDDI------- 400

Db        392 VNNYRSTKYEMYSKNAEKPSRSKRTIKQKQRRKFMAKPAEEQLDVGQSKDENIHTSHITQ 451

Qy        401 ------SDTFSEPGYE--NDSVEDLKEVTSISSRKRG---------KRRYFWEYSEQLTP 443

Db        452 DEFQRNSDRNMEEHEEMGNDCVSK-KQMPPVGSKKSSTRKDKEESKKKRFSSESKNKLVP 510

Qy        444 SQ------QERML--RPSEW-----NRDTLPSNMYQKNGL--HHGKYAVKKSRRTDVEDL 488

Db        511 EEVTSTVTKSRRISRRPSDWWVVKSEESPVYSNSSVRNELPMHH-----NSSRK------ 559

Qy        489 TPNPKKLLQIGNELRK 504

Db        560 --STKKTNQSSKNIRK 573


RESULT 12
US-11-054-281-107
; Sequence 107, Application US/11054281
; Publication No. US20060013813A1
; GENERAL INFORMATION:
;  APPLICANT: Mezes et al.
;  TITLE OF INVENTION: Proteins and Nucleic Acids Encoding Same
;  FILE REFERENCE: 21402-240CIP
;  CURRENT APPLICATION NUMBER: US/11/054,281
;  CURRENT FILING DATE:  2005-02-08
;  PRIOR APPLICATION NUMBER: 60/261,014
;  PRIOR FILING DATE: 2001-01-11
;  PRIOR APPLICATION NUMBER: 60/261,018
;  PRIOR FILING DATE: 2001-01-11
;  PRIOR APPLICATION NUMBER: 60/318,410
;  PRIOR FILING DATE: 2001-09-10
;  PRIOR APPLICATION NUMBER: 60/261,013
;  PRIOR FILING DATE: 2001-01-11
;  PRIOR APPLICATION NUMBER: 60/261,026
;  PRIOR FILING DATE: 2001-01-11
;  PRIOR APPLICATION NUMBER: 60/261,029
;  PRIOR FILING DATE: 2001-01-11
;  PRIOR APPLICATION NUMBER: 60/313,170
;  PRIOR FILING DATE: 2001-08-17
;  PRIOR APPLICATION NUMBER: 10/044,564
;  PRIOR FILING DATE: 2002-01-11
;  NUMBER OF SEQ ID NOS: 324
;  SOFTWARE: PatentIn Ver. 2.1
; SEQ ID NO 107
;   LENGTH: 1229
;   TYPE: PRT
;   ORGANISM: Rattus norvegicus
US-11-054-281-107
```

```
  Query Match             4.2%;  Score 142;  DB 7;  Length 1229;
  Best Local Similarity   24.7%;  Pred. No. 0.16;
  Matches   67;  Conservative   47;  Mismatches   83;  Indels   74;  Gaps   16;

Qy        250 PLLSQIHTDAAKENTCYC------------GAVAK----RQEKKGMEPLQGHATPALPF 292
              | |:    : || :|| |              | ||:      |::|   |    |
Db         45 PTLTMPGRNKAK-STCSCPDLQPNGQDLGESGRVARLGADESEEEG-RSLSNVGDP---- 98

Qy        293 KETQELLLSPLP---------QEGPGSLAAGESSSLSASTSVSDSSQKKEEHNYSLFVSD 343
                |:: ||           | |  :|   : |::   |  :   :::||        :
Db         99 ----EIIKSPSDPKQYRYIKLQNGLQALLISDLSNVEGKTGNATDEEEEEEE------EE 148

Qy        344 NLGEQPTKCSPEEDEEDEEDVDDEDHDEGFGSEHELSENEEEE-EEEEDYEDDKDDDISD 402
                ||       ||:||:|:| ||:| | |     |: :::||  ::||:::||: ||  |
Db        149 EEGE-------EEEEEEEDDDDDDDEDSG----AEIQDDDEEGFDDEEEFDDDEHDD--D 195

Qy        403 TFSEPGYENDSVEDLKEVTSISSRKRGKRRYFWEYSEQLTPSQQERMLRPSEWNR--DTL 460
                  | : :|:  |    :: |:  :  |  :| :||    :|:   | :::   ||
Db        196 DLDNEENELEELEERVEARKKTTEKQQSQNLFLLWS-KLT----DRLWFKSSYSKMSSTL 250

Qy        461 ---PSNMYQKNGLHHGKYAVKKSRRTDVEDL 488
                  |:|   |         :||   || |
Db        251 LVETRNLYGVVG--------AESRSAPVEHL 273


RESULT 13
US-10-689-742-142
; Sequence 142, Application US/10689742
; Publication No. US20050250180A1
; GENERAL INFORMATION:
;  APPLICANT: Jacobs, Kenneth
;  APPLICANT:  McCoy, John  M
;  APPLICANT:  LaVallie, Edward R
;  APPLICANT:  Racie, Lisa A
;  APPLICANT:  Evans, Cheryl
;  APPLICANT:  Merberg, David
;  APPLICANT:  Treacy, Maurice
;  APPLICANT:  Spaulding, Vikki
;  TITLE OF INVENTION: SECRETED PROTEINS AND POLYNUCLEOTIDES ENCODING THEM
;  FILE REFERENCE: 00766.000091.10
;  CURRENT APPLICATION NUMBER: US/10/689,742
;  CURRENT FILING DATE:  2003-10-22
;  PRIOR APPLICATION NUMBER: 09/746,783
;  PRIOR FILING DATE: 2000-12-21
;  NUMBER OF SEQ ID NOS: 231
;  SOFTWARE: PatentIn version 3.2
; SEQ ID NO 142
;   LENGTH: 903
;   TYPE: PRT
;   ORGANISM: Homo sapiens
US-10-689-742-142

  Query Match             4.2%;  Score 140.5;  DB 6;  Length 903;
  Best Local Similarity   20.2%;  Pred. No. 0.14;
  Matches  142;  Conservative  109;  Mismatches  227;  Indels  225;  Gaps   38;

Qy         35 NSSDPDFMYELDREMNYQQ-----NPRD--------NFLSLEDC--KDIENLESFTDVLD 79
              |||: |   : |:|   |:      | |           | |:||  ::|: :      :
Db        129 NSSEEDSPIKSDKESVEQRSTVVDNDADFQGTKRACRCLILDDCEKREIKKVN-----VS 183

Qy         80 NEGALTSNWEQWDTYCEDLTKYTKLTSCDIWGTKEVDYLGLDDFSSPYQDEEVISKTPTL 139
              || | |          |::| |  :                |:||  |
Db        184 EEGPLNS------AVVEEITGYLAVN-------------GVDDSDS-------------- 210

Qy        140 AQLNSEDSQSVSDSLYYPDSLFSVKQNPLPSSFPGKKITSRAAAPVCSSKTLQAEVPLSD 199
              | :| :| |          ||   : ||| : |              |   |::
Db        211 AVINCDDCQ--------PDG--NTKQNSIGSY-------------VLQEKSV-------- 239

Qy        200 CVQKASKPPSSTQIMVKTNMYHNEK-----VNFHVECKDYVKKAKVKIN---------PV 245
                    |    : ||    |:|: : :     ::  | | |   : :||:          ||
Db        240 ----AENGDTDTQ----TSMFLDSRKEDSYIDHKVPCTD--SQVQVKLEDHKIVTACLPV 289

Qy        246 QQSRPLLSQIHT-----------DAAKENTCYCGAVAKRQEKK---------GME--PLQ 283
```

```
                :    | ::  |            |: :|      : | ::: |          :|  |
Db        290 EHVNQLTTEPATGPFSETQSSLRDSEEEVDVVGDSSASKEQCKENTNNELDTSLESMPAS 349

Qy        284 G--HATPALPFKETQELLLSPLPQEGPGSL------AAGESSSLSASTSVSDSSQKKEEH 335
              |    :| |      | : ||  |||   :|      ||    | : ::|    : : ||:
Db        350 GEPEPSPVLDCVSAQMMSLSE-PQEHRYTLRTSPRRAAPTRGSPTKNSSPYRENGQFEEN 408

Qy        336 NYS-----LFVSDNLGEQPTKCSPEEDEEDEEDV--DDEDH-DEGFG---SEHELSENE- 383
              | |        ||||: : ||  :| |  ::|: :  | :::  ||      ||  |:
Db        409 NLSPNETNATVSDNVSQSPT--NPGEISQNEKGICCDSQNNGSEGVSKPPSEARLNIGHL 466

Qy        384 -EEEEEEEDYEDDKDDDISDTFSEPGYENDSV-----EDLKEVTSISSRKRGKRRYFWEY 437
                 :|    :  :::||  | :    :|:| |      :| : :    :    :|   :
Db        467 PSAKESASQHITEEEDDDPDVYY---FESDHVALKHNKDYQRLLQTIAVLEAQRSQAVQD 523

Qy        438 SEQLTPSQQERMLRPSEWNRDTLPSNMYQKNGLHHGKYAVKKSRRTDVEDLTPNPKKLLQ 497
              | |   |:| :  |                    :  |  :: |:    | |::::|
Db        524 LESLGRHQREALKNPI-------------------GFVEKLQKKADIG--LPYPQRVVQ 561

Qy        498 ------------IGNELRKLN---------KVISDLTPVSELPLTARPRS----RKEKNK 532
                         :||  |:           |:: |     :: | |||:|    :|:
Db        562 LPEIVWDQYTHSLGNFEREFKNRKRHTRRVKLVFD-----KVGLPARPKSPLDPKKDGES 616

Qy        533 LAFRACRLKKKAQYEANKVKLW------GLNTEYDNLLFVINSIKQEIVNRVQNPRDERG 586
              |::    |   :  ::: ::          |  | |: :   |:     :: | :|
Db        617 LSYSMLPLSDGPEGSSSRPQMIRGRLCDDTKPETFNQLWTVEEQKKLEQLLIKYPPEEVE 676

Qy        587 PNMGQKLEILIKDTLGLPVAGQTSEFVNQVLEK-TAEGNPTGG 628
                  ||    | | ||   | | :  |  :   | |  | | |
Db        677 SRRWQK----IADELGNRTAKQVASRVQKYFIKLTKAGIPVPG 715


RESULT 14
US-11-124-368A-305
; Sequence 305, Application US/11124368A
; Publication No. US20050287559A1
; GENERAL INFORMATION:
;  APPLICANT: Michele Cargill
;  APPLICANT:  James J. Devlin
;  APPLICANT:  May Luke
;  TITLE OF INVENTION: Genetic Polymorphisms Associated with
;  TITLE OF INVENTION:  Vascular Diseases, Methods of Detection and Uses Thereof
;  FILE REFERENCE: CL001524
;  CURRENT APPLICATION NUMBER: US/11/124,368A
;  CURRENT FILING DATE:  2005-05-09
;  PRIOR APPLICATION NUMBER: US 60/568,845
;  PRIOR FILING DATE: 2004-05-07
;  PRIOR APPLICATION NUMBER: US 60/625,936
;  PRIOR FILING DATE: 2004-11-09
;  NUMBER OF SEQ ID NOS: 21112
;  SOFTWARE: FastSEQ for Windows Version 4.0
; SEQ ID NO 305
;   LENGTH: 2801
;   TYPE: PRT
;   ORGANISM: Homo sapiens
US-11-124-368A-305

  Query Match             4.2%;  Score 140.5;  DB 7;  Length 2801;
  Best Local Similarity   19.9%;  Pred. No. 0.56;
  Matches  131;  Conservative   89;  Mismatches  242;  Indels  195;  Gaps   27;

Qy          2 PQPSVSGMDPPFGDAFRSHTFSEQTLMSTDLLANSSDPDFMYELDREMNYQQNPRDNFLS 61
              |:|:||        |    :|| |  :   ||| |          |       | |::   :
Db       1935 PKPAVS-------DEKNINTFVETPVQKLDLLGN---------LPGSKRQPQTPKEKAEA 1978

Qy         62 LEDCKDIENL--------ESFTDVLDNEGALTS-NWEQWDT----------------YCE 96
              |||    : |        || ||    | :  |   | : |                |
Db       1979 LEDLVGFKELFQTPGHTEESMTDDKITEVSCKSPQPESFKTSRSSKQRLKIPLVKVDMKE 2038

Qy         97 DLTKYTKLT----------------SCDIWGTKEVDYLGLDDFSSPYQD-EEVISKTPTL 139
              :   :|||                | |   ||          || |  :  | | :
Db       2039 EPLAVSKLTRTSGETTQTHTEPTGDSKSIKAFKE----------SPKQILDPAASVTGSR 2088
```

```
Qy        140 AQLNS--EDSQSVSDSLYYPDSLFSV----------KQNPLPSSFPGKKITSRAAAPVC 186
              || :  | :::: | : : : |||             |   :|   |  ::|  |
Db       2089 RQLRTRKEKARALEDLVDFKE-LFSAPGHTEESMTIDKNTKIPCKSPPPELTDTAT---- 2143

Qy        187 SSKTLQAEVPLSDCVQKASKPPSSTQIMVKTNMYHNEKVNFHVECKDYVKKAKVKINPVQ 246
              |:|      |  :  :: |     ||   ::   | |  :     |  ::|| | |||:
Db       2144 STKRCPKTRPRKEVKEELSAVERLTQTSGQSTHTHKEPASGDEGIKVLKQRAKKKPNPVE 2203

Qy        247 QS----RPLLSQIHTDAAKENTCYCGAVAKRQEKKGMEPLQGHATPALPFKETQELLLSP 302
              :    ||                   | ::: : :| | |    :     |||
Db       2204 EEPSRRRPR-----------------APKEKAQPLEDLAGFTELSETSGHTQE----- 2239

Qy        303 LPQEGPGSLAAGESSSL---SASTSVSDSSQKKEEHNYSLFVSDNLGEQP-----TKCSP 354
              || ||::: :   |     | |::   : |  :      : |:|     |: |
Db       2240 -------SLTAGKATKIPCESPPLEVVDTTASTKRHLRTRVQKVQVKEEPSAVKFTQTSG 2292

Qy        355 EEDEEDEEDVDDEDHDEGF-----------------GSEHELSENEEEEEEEDYEDDK 396
              |  : |:|   :   |:|                 ||    |  :  ||    |
Db       2293 ETTDADKEPAGE---DKGIKALKESAKQTPAPAASVTGSRRRPRAPRESAQAIEDLAGFK 2349

Qy        397 DDDISDTFSEPGYENDSVEDLK-----------EVTSISSRKRGKRRYFWEYSEQLTPS 444
              |          |:  :|: | |            | |: ||::| : |     ::::
Db       2350 D-------PAAGHTEESMTDDKTTKIPCKSSPELEDTATSSKRRPRTR-----AQKV--E 2395

Qy        445 QQERMLRPSEWNRDTLPSNMYQKNGLHHGK----YAVKKSRRTDVEDLTPN------PKK 494
              :| :|   :  :  :   | :  ||    :     |: | ||:  :      ||:
Db       2396 VKEELLAVGKLTQTSGETTHTDKEPVGEGKGTKAFKQPAKRKLDAEDVIGSRRQPRAPKE 2455

Qy        495 LLQIGNELRKLNKVISDLTPVSEL------PLTARPRSRKEKN---KLAFRACRLKK 542
                |  :|   ::       ||      |: |:   :     |:: |  |  |
Db       2456 KAQPLEDLASFQELSQTPGHTEELANGAADSFTSAPKQTPDSGKPLKISRRVLRAPK 2512


RESULT 15
US-11-124-367A-433
; Sequence 433, Application US/11124367A
; Publication No. US20060024700A1
; GENERAL INFORMATION:
;  APPLICANT: Michele Cargill
;  APPLICANT:  Hongjin Huang
;  TITLE OF INVENTION: Genetic Polymorphisms Associated with
;  TITLE OF INVENTION:  Fibrosis Methods of Detection and Uses Thereof
;  FILE REFERENCE: CL001519.ORD
;  CURRENT APPLICATION NUMBER: US/11/124,367A
;  CURRENT FILING DATE:  2005-05-09
;  PRIOR APPLICATION NUMBER: US 60/568,846
;  PRIOR FILING DATE: 2004-05-07
;  PRIOR APPLICATION NUMBER: US 60/582,609
;  PRIOR FILING DATE: 2004-06-25
;  PRIOR APPLICATION NUMBER: US 60/599,554
;  PRIOR FILING DATE: 2004-08-09
;  NUMBER OF SEQ ID NOS: 34460
;  SOFTWARE: FastSEQ for Windows Version 4.0
; SEQ ID NO 433
;   LENGTH: 2801
;   TYPE: PRT
;   ORGANISM: Homo sapiens
US-11-124-367A-433

  Query Match             4.2%;  Score 140.5;  DB 7;  Length 2801;
  Best Local Similarity   19.9%;  Pred. No. 0.56;
  Matches  131;  Conservative   89;  Mismatches  242;  Indels  195;  Gaps   27;

Qy          2 PQPSVSGMDPPFGDAFRSHTFSEQTLMSTDLLANSSDPDFMYELDREMNYQQNPRDNFLS 61
              |:|:||      |    :|| |  :   ||| |          |       | |::   :
Db       1935 PKPAVS-------DEKNINTFVETPVQKLDLLGN---------LPGSKRQPQTPKEKAEA 1978

Qy         62 LEDCKDIENL--------ESFTDVLDNEGALTS-NWEQWDT----------------YCE 96
              |||    : |       || ||    | :  |   | : |                  |
Db       1979 LEDLVGFKELFQTPGHTEESMTDDKITEVSCKSPQPESFKTSRSSKQRLKIPLVKVDMKE 2038

Qy         97 DLTKYTKLT----------------SCDIWGTKEVDYLGLDDFSSPYQD-EEVISKTPTL 139
              :    :|||                |  |   ||          || |  :   | | :
Db       2039 EPLAVSKLTRTSGETTQTHTEPTGDSKSIKAFKE----------SPKQILDPAASVTGSR 2088

Qy        140 AQLNS--EDSQSVSDSLYYPDSLFSV----------KQNPLPSSFPGKKITSRAAAPVC 186
              || :  | :::: | : : : |||             |   :|   |  ::|  |
Db       2089 RQLRTRKEKARALEDLVDFKE-LFSAPGHTEESMTIDKNTKIPCKSPPPELTDTAT---- 2143

Qy        187 SSKTLQAEVPLSDCVQKASKPPSSTQIMVKTNMYHNEKVNFHVECKDYVKKAKVKINPVQ 246
              |:|      |  :  :: |     ||   ::   | |  :     |  ::|| | |||:
Db       2144 STKRCPKTRPRKEVKEELSAVERLTQTSGQSTHTHKEPASGDEGIKVLKQRAKKKPNPVE 2203

Qy        247 QS----RPLLSQIHTDAAKENTCYCGAVAKRQEKKGMEPLQGHATPALPFKETQELLLSP 302
              :    ||                   | ::: : :| | |    :     |||
Db       2204 EEPSRRRPR-----------------APKEKAQPLEDLAGFTELSETSGHTQE----- 2239

Qy        303 LPQEGPGSLAAGESSSL---SASTSVSDSSQKKEEHNYSLFVSDNLGEQP-----TKCSP 354
              || ||::: :   |     | |::   : |  :      : |:|     |: |
Db       2240 -------SLTAGKATKIPCESPPLEVVDTTASTKRHLRTRVQKVQVKEEPSAVKFTQTSG 2292

Qy        355 EEDEEDEEDVDDEDHDEGF-----------------GSEHELSENEEEEEEEDYEDDK 396
              |  : |:|   :   |:|                 ||    |  :  ||    |
Db       2293 ETTDADKEPAGE---DKGIKALKESAKQTPAPAASVTGSRRRPRAPRESAQAIEDLAGFK 2349

Qy        397 DDDISDTFSEPGYENDSVEDLK-----------EVTSISSRKRGKRRYFWEYSEQLTPS 444
              |          |:  :|: | |            | |: ||::| : |     ::::
Db       2350 D-------PAAGHTEESMTDDKTTKIPCKSSPELEDTATSSKRRPRTR-----AQKV--E 2395

Qy        445 QQERMLRPSEWNRDTLPSNMYQKNGLHHGK----YAVKKSRRTDVEDLTPN------PKK 494
              :| :|   :  :  :   | :  ||    :     |: | ||:  :      ||:
Db       2396 VKEELLAVGKLTQTSGETTHTDKEPVGEGKGTKAFKQPAKRKLDAEDVIGSRRQPRAPKE 2455

Qy        495 LLQIGNELRKLNKVISDLTPVSEL------PLTARPRSRKEKN---KLAFRACRLKK 542
                |  :|   ::       ||      |: |:   :     |:: |  |  |
Db       2456 KAQPLEDLASFQELSQTPGHTEELANGAADSFTSAPKQTPDSGKPLKISRRVLRAPK 2512


Search completed: February 28, 2006, 09:02:38
Job time : 19.0827 secs
```

OM protein - protein search, using sw model

Run on: February 28, 2006, 08:45:35 ; Search time 38.9699 Seconds
(without alignments)
1577.691 Million cell updates/sec

Title: US-10-717-665A-44
Perfect score: 3349
Sequence: 1 MPQPSVSGMDPPFGDAFRSH.....TAEGNPTGGLVGLRIPTSKV 639

Scoring table: BLOSUM62
Gapop 10.0 , Gapext 0.5

Searched: 283416 seqs, 96216763 residues

Total number of hits satisfying chosen parameters: 283416

Minimum DB seq length: 0
Maximum DB seq length: 2000000000

Post-processing: Minimum Match 0%
Maximum Match 100%
Listing first 45 summaries

Database : PIR_80:*
1: pir1:*
2: pir2:*
3: pir3:*
4: pir4:*

Pred. No. is the number of results predicted by chance to have a score greater than or equal to the score of the result being printed, and is derived by analysis of the total score distribution.

SUMMARIES

Result No.	Score	% Query Match	Length	DB	ID	Description
1	209	6.2	707	1	DNMS	nucleolin - mouse
2	202	6.0	712	2	JH0148	nucleolin - rat
3	188	5.6	2761	2	T29285	hypothetical prote
4	184	5.5	1089	2	S48244	NMD2 protein - yea
5	180.5	5.4	694	1	DNCHNL	nucleolin - chicke
6	179	5.3	990	2	T38274	probable transcrip
7	174	5.2	707	2	A35804	nucleolin - human
8	168.5	5.0	1187	2	T46637	transcription fact
9	166.5	5.0	412	2	A55320	immunophilin FKBP4
10	165.5	4.9	1063	2	A40253	acidic nuclear pro
11	165	4.9	678	2	A54514	glutamic acid-rich
12	165	4.9	1038	2	S52522	hypothetical prote
13	164.5	4.9	1082	2	T50650	AP-3 complex beta3
14	164	4.9	916	2	S22864	DNA topoisomerase
15	163.5	4.9	802	1	S48529	NAB3 protein - yea
16	162.5	4.9	713	2	A27441	nucleolin - Chines
17	162.5	4.9	836	2	S49940	cell division cont
18	161	4.8	831	2	S39835	hypothetical prote
19	161	4.8	885	2	G71608	ATP-dept. acyl-CoA
20	160.5	4.8	1877	2	T21861	hypothetical prote
21	160.5	4.8	1974	2	T16703	hypothetical prote
22	160	4.8	500	2	S55785	nucleolar protein
23	160	4.8	1188	2	T46608	zinc finger protei
24	159.5	4.8	1128	2	G86266	hypothetical prote
25	159	4.7	736	2	G01522	acidic 82 kDa prot
26	159	4.7	1085	2	S55352	IFH1 protein - yea
27	159	4.7	1115	2	T41342	probable coiled-co
28	158.5	4.7	1611	2	T38236	hypothetical prote
29	158.5	4.7	3488	2	T34418	hypothetical prote
30	158	4.7	1969	2	T38495	hypothetical prote
31	157.5	4.7	1151	2	T24541	hypothetical prote
32	157	4.7	562	2	S38149	SIS2 protein - yea
33	156.5	4.7	1875	2	S38173	myosin-like protei
34	156	4.7	695	1	A49795	Alzheimer's diseas
35	156	4.7	1726	2	A45948	major merozoite su
36	155.5	4.6	1280	2	T00365	hypothetical prote
37	155	4.6	699	2	T01029	hypothetical prote
38	155	4.6	740	2	T03847	Fas-binding protei
39	155	4.6	1435	2	A37793	erythrocyte-bindin
40	154.5	4.6	1182	2	T30189	myelin transcripti
41	154.5	4.6	1190	2	S47536	oxysterol-binding
42	154	4.6	1804	2	T34518	nestin - golden ha
43	153.5	4.6	992	2	T38817	hypothetical prote
44	153	4.6	610	2	S67701	hypothetical prote
45	153	4.6	792	2	T42963	hypothetical prote

ALIGNMENTS

RESULT 1
DNMS
nucleolin - mouse
C;Species: Mus musculus (house mouse)
C;Date: 30-Sep-1992 #sequence_revision 30-Sep-1992 #text_change 09-Jul-2004
C;Accession: A29958; A40769; A56240; I84688
R;Bourbon, H.M.; Lapeyre, B.; Amalric, F.
J. Mol. Biol. 200, 627-638, 1988
A;Title: Structure of the mouse nucleolin gene. The complete sequence reveals that each F
A;Reference number: A29958; MUID:88316930; PMID:3137346
A;Accession: A29958
A;Molecule type: DNA
A;Residues: 1-707 <BOU>
A;Cross-references: UNIPROT:P09405; UNIPARC:UPI00000009C3; GB:X07699; NID:g53453; PIDN:C
R;Pasternack, M.S.; Bleier, K.J.; McInerney, T.N.
J. Biol. Chem. 266, 14703-14708, 1991
A;Title: Granzyme A binding to target cell proteins. Granzyme A binds to and cleaves nucl
A;Reference number: A40769; MUID:91317840; PMID:1860869
A;Accession: A40769
A;Molecule type: protein
A;Residues: 2-20,'X',22-24 <PAS>
A;Cross-references: UNIPARC:UPI00001739CF
R;Yang, T.H.; Tsai, W.H.; Lee, Y.M.; Lei, H.Y.; Lai, M.Y.; Chen, D.S.; Yeh, N.H.; Lee, S.
Mol. Cell. Biol. 14, 6068-6074, 1994
A;Title: Purification and characterization of nucleolin and its identification as a trans
A;Reference number: A56240; MUID:94344117; PMID:8065340
A;Accession: A56240
A;Molecule type: protein
A;Residues: 2-19;558-567 <YAN>
A;Cross-references: UNIPARC:UPI00001739D0; UNIPARC:UPI00001739D1
R;Bourbon, H.
Gene 68, 73-84, 1988
A;Title: Sequence and structure of the nucleolin promoter in rodents: Characterization of
A;Reference number: I48118; MUID:89121496; PMID:2906027
A;Accession: I84688
A;Status: preliminary; translated from GB/EMBL/DDBJ
A;Molecule type: DNA
A;Residues: 1-44 <RES>
A;Cross-references: UNIPARC:UPI000016CF68; GB:M22089; NID:g200111; PIDN:AAA39841.1; PID:g
C;Comment: This housekeeping protein is involved in the synthesis, packaging, and maturat
C;Genetics:
A;Introns: 6/3; 45/3; 211/1; 273/1; 302/1; 349/2; 391/1; 432/3; 484/1; 525/2; 566/1; 608/
C;Superfamily: nucleolin; ribonucleoprotein repeat homology
C;Keywords: DNA binding; duplication; nucleus; phosphoprotein; RNA binding; transcriptior
F;310-375/Domain: ribonucleoprotein repeat homology <RRM1>
F;311-316/Region: RNA-binding RNP2 motif
F;349-356/Region: RNA-binding RNP1 motif
F;396-458/Domain: ribonucleoprotein repeat homology <RRM2>
F;397-402/Region: RNA-binding RNP2 motif
F;431-438/Region: RNA-binding RNP1 motif
F;488-551/Domain: ribonucleoprotein repeat homology <RRM3>
F;489-494/Region: RNA-binding RNP2 motif

```
F;524-531/Region: RNA-binding RNP1 motif
F;570-634/Domain: ribonucleoprotein repeat homology <RRM4>
F;571-576/Region: RNA-binding RNP2 motif
F;607-614/Region: RNA-binding RNP1 motif

  Query Match             6.2%;  Score 209;  DB 1;  Length 707;
  Best Local Similarity   21.0%;  Pred. No. 0.0012;
  Matches  128;  Conservative   78;  Mismatches  180;  Indels  224;  Gaps   28;

Qy        137 PTLAQLNSEDSQSVSDSLYYPDSLFSVKQNPLPSSFPGKKITSRAAAPVCSSKTLQAEVP 196
              |   : :||| :   |     |     ::  :|    ||| |:  |  |  |:| :| ||
Db         21 PKEVEEDSEDEEMSEDE----DDSSGEEEVVIPQK-KGKKATTTPAKKVVVSQTKKAAVP 75

Qy        197 LSDCVQKASKPPSSTQIMVKTNMYHNEKVNFHVECKDYVKKAKVKINPVQQSRPLLSQIH 256
                :||:  |                          |     ||  | | :
Db         76 TP--AKKAAVTPGK--------------------KAVATPAKKNITPAK---------- 102

Qy        257 TDAAKENTCYCGAVAKRQEKKGMEPLQGHATPALPFKETQELLLSPLP-QEGPGSLAAGE 315
                      |     |||                 |   | | | ::|  : | |
Db        103 -------------VIPTPGKKG--------------AAQAKALVPTPGKKGAATPAKGA 134

Qy        316 SSSLSASTSVSDSSQKKEEHNYSLFVSDN----------LGEQPTKCSP----EEDEED 360
               :  :|    ||  : :|: : |    |:             | :| | :|     |||||
Db        135 KNGKNAKKEDSDEDEDEEDEDDSDEDEDDEEEDEFEPPIVKGVKPAKAAPAAPASEDEED 194

Qy        361 EEDVDDEDHDEGFGSEHELSENEEE----------------------EEEEEDYEDDK 396
              :|| |||: |:    | |  ::|||                        |||:| |:|:
Db        195 DEDEDDEEDDD----EEEEDDSEEEVMEITTAKGKKTPAKVVPMKAKSVAEEEDDEEEDE 250

Qy        397 DDDISDTFSEPGYENDSVEDLKEVTSISSRKRGKRRYFWEYSEQL-TPSQQERMLRPSEW 455
              ||:  |   |   ::|  |: :|    :  || |     | ::|   |  ::: :  ||
Db        251 DDEDEDDEEEDDEDDDEEEEEEEPVKAAPGKRKK-----EMTKQKEAPEAKKQKVEGSE- 304

Qy        456 NRDTLPSNMYQKN-----GLHHGKYAVKK---------------SRR------TDVEDL 488
                | | |::  |      ::  |:|: :                :|:          |||
Db        305 --PTTPFNLFIGNLNPNKSVNELKFAISELFAKNDLAVVDVRTGTNRKFGYVDFESAEDL 362

Qy        489 TPNPKKLLQI------GNELRKLNKVISDLTPVSELPLTARPRSRKEKNKLAFRACRLK- 541
                :| |::      |||: || |             |: |  |   | |   |
Db        363 ----EKALELTGLKVFGNEI-KLEK---------------PKGRDSKKVRAARTLLAKN 401

Qy        542 ----------KKAQYEANKVKLWGLNTEYDNLLFVINSIKQEIVNRVQNPRDERGPNM-- 589
                       |:   :| :::|   : :   : ::    | |  :  :|  :::|  :
Db        402 LSFNITEDELKEVFEDAMEIRLVSQDGKSKGIAYI--EFKSE-ADAEKNLEEKQGAEIDG 458

Qy        590 -----------GQKLE-------------ILIKDTLGLPVAGQTSEFVNQVLEKTA---- 621
                        ||: |              |:   |       | | : :| ||
Db        459 RSVSLYYTGEKGQRQERTGKTSTWSGESKTLVLSNLSY---SATKETLEEVFEKATFIKV 515

Qy        622 ----EGNPTG 627
                  | | |
Db        516 PQNPHGKPKG 525


RESULT 2
JH0148
nucleolin - rat
C;Species: Rattus norvegicus (Norway rat)
C;Date: 31-Dec-1991 #sequence_revision 31-Dec-1991 #text_change 23-Jul-1999
C;Accession: JH0148; A24088; I63130
R;Bourbon, H.M.; Amalric, F.
Gene 88, 187-196, 1990
A;Title: Nucleolin gene organization in rodents: highly conserved sequences within three
A;Reference number: JH0148; MUID:90269607; PMID:2347493
A;Accession: JH0148
A;Molecule type: DNA
A;Residues: 1-712 <BOU>
A;Cross-references: UNIPARC:UPI000017723A
A;Note: the authors translated the initiation codon GTG for residue 1 as Val
R;Lischwe, M.A.; Cook, R.G.; Ahn, Y.S.; Yeoman, L.C.; Busch, H.
Biochemistry 24, 6025-6028, 1985
A;Title: Clustering of glycine and Ng,Ng-dimethylarginine in nucleolar protein C23.
A;Reference number: A24088; MUID:86104094; PMID:4084504
A;Accession: A24088
A;Molecule type: protein
A;Residues: 651-703 <LIS>
A;Cross-references: UNIPARC:UPI000017723B
R;Bourbon, H.
Gene 68, 73-84, 1988
A;Title: Sequence and structure of the nucleolin promoter in rodents: Characterization of
A;Reference number: I48118; MUID:89121496; PMID:2906027
A;Accession: I63130
A;Status: preliminary; translated from GB/EMBL/DDBJ
A;Molecule type: DNA
A;Residues: 'MV',2-44 <RES>
A;Cross-references: UNIPARC:UPI0000170A72; GB:M22090; NID:g205793; PIDN:AAA41733.1; PID:g
C;Comment: This protein is the major nucleolar-specific protein in eukaryotic exponential
C;Genetics:
A;Gene: nuc
A;Start codon: GTG
A;Introns: 5/3; 44/3; 211/1; 274/1; 303/1; 350/2; 392/1; 433/2; 485/1; 526/2; 571/1; 613/
C;Superfamily: nucleolin; ribonucleoprotein repeat homology
C;Keywords: DNA binding; nucleus
F;311-376/Domain: ribonucleoprotein repeat homology <RRM1>
F;397-459/Domain: ribonucleoprotein repeat homology <RRM2>
F;489-552/Domain: ribonucleoprotein repeat homology <RRM3>
F;575-639/Domain: ribonucleoprotein repeat homology <RRM4>

  Query Match             6.0%;  Score 202;  DB 2;  Length 712;
  Best Local Similarity   22.4%;  Pred. No. 0.0027;
  Matches  107;  Conservative   60;  Mismatches  171;  Indels  140;  Gaps   18;

Qy        129 DEEVISKTPTLAQLNSEDSQSVSDSLYYPDSLFSVKQNPLPSSFPGKKITSRAAAPVCSS 188
              : : ::  |   : :||| :   |     |      ::  :|    ||| |:  |  |  |
Db         12 ESKKMAPPPKEVEEDSEDEEMSEDE----DDSSGEEEVVIPQK-KGKKATTTPAKKVVVS 66

Qy        189 KTLQAEVPLSDCVQKASKPPSSTQIMVKTNMYHNEKVNFHVECKDYVKKAKVKINPVQQS 248
              :| :| ||     :||:  |                |       |  |  |||   |
Db         67 QTKKAAVPTP--AKKAAVTPG-------------KKAAATPAKKAVTPAKVVPTP---- 106

Qy        249 RPLLSQIHTDAAKENTCYCGAVAKRQEKKGMEPLQGHATPALPFKETQELLLSPLPQEGP 308
                        |:      |:     |||       |||    |
Db        107 -----------GKKGAAQAKALVPTPGKKGA------VTPAKGAKN-------------- 135

Qy        309 GSLAAGESSSLSASTSVSDSSQKKEEHNYSLFVSDNLGEQPTKCSP------EEDEEDEE 362
              |  |  | |         | | : |:           | :| | :|      :|||||::
Db        136 GKNAKKEDSDEDEDEEDEDDSDEDEDEEDEFEPPVVKGVKPAKAAPAAPASEDEDEEDDD 195

Qy        363 DVDDEDHDEGFGSEHELSENEEE----------------------EEEEEDYEDDKDD 398
              | ||:| ||    | |  ::|||                       |||||  |||:|:
Db        196 DEDDDDDDE---EEEEEDDSEEEVMEITPAKGKKTPAKVVPVKAKSVAEEEEDDEDDEDE 252

Qy        399 DISDTFSEPGYENDSVEDLKEVTSISSRKRGKRRYFWEYSEQLTPSQQ--ERMLRPSEWN 456
              :  :   |   |:|  || :|       |||:      :::|  ::  |   :  | :
Db        253 EEDE--DEEDEEDDEDEDEEEEEEPVKAAPGKRK-----KEMTKQKEAPEAKKQKIEGS 304

Qy        457 RDTLPSNMYQKNGLHHGKYAVKKSRRTDVEDLTPNPKKLLQIGNELRKLNKVISDLTPVS 516
                | | |::                  : :| ||          : :|   ||:|   :
Db        305 EPTTPFNLF------------------IGNLNPN--------KSVAELKVAISELFAKN 337

Qy        517 EL-PLTARPRSRKEKNKLAFRACRLKKKAQYEANKVKLWGLNTEYDNLLFVINSIKQE 573
              :|  :  |  : ::   : | :    :||  |   :|::|            | || |
Db        338 DLAAVDVRTGTNRKFGYVDFESAEDLEKA-LELTGLKVFG-----------NEIKLE 382


RESULT 3
T29285
hypothetical protein C34D4.14 - Caenorhabditis elegans
C;Species: Caenorhabditis elegans
C;Date: 15-Oct-1999 #sequence_revision 15-Oct-1999 #text_change 09-Jul-2004
C;Accession: T29285
R;Du, Z.; Le, T.T.
submitted to the EMBL Data Library, May 1996
A;Description: The sequence of C. elegans cosmid C34D4.
A;Reference number: Z20600
A;Accession: T29285
```

```
A;Status: preliminary; translated from GB/EMBL/DDBJ
A;Molecule type: DNA
A;Residues: 1-2761 <DUZ>
A;Cross-references: UNIPROT:Q18447; UNIPARC:UPI000007762A; EMBL:U58755; PIDN:AAB00699.1;
A;Experimental source: strain Bristol N2; clone C34D4
C;Genetics:
A;Gene: CESP:C34D4.14
A;Map position: 4
A;Introns: 120/1; 201/3; 365/3; 445/2; 509/1; 728/3; 856/3; 902/3; 940/3; 1027/3; 1508/3

  Query Match             5.6%;  Score 188;  DB 2;  Length 2761;
  Best Local Similarity   22.3%;  Pred. No. 0.076;
  Matches   88;  Conservative   56;  Mismatches  131;  Indels  120;  Gaps   15;

Qy        170 SSFPGKKI----TSRAAAPVCSSKTLQAEVPLSDCVQKASKPPSSTQIMVKTNMYHNEKV 225
              | ||   :    :|:: :|  :|: :     :|        ||||:      :      :
Db       1609 SPFPNLPVPPWRSSKSTSPAIASRLINTTTTVSGTSSPPPPPPSSSTFSSLASGL-GFGL 1667

Qy        226 NFHVECKDYVKKAKVKINPVQQSRPLLSQIHTDAAKENTCYCGA---------------- 269
              | |                  | ::|  | :   |: :||   |
Db       1668 NRH---------------KQHNKPAASALSRFASVKNTTPAGTPSSGGSSGGAIGKKSM 1711

Qy        270 -----VAKRQEKKG-----------MEPLQGHATPALPFKETQELLLSPLP--------- 304
                  | :||:  |             | || | ||:|     : ||   :|
Db       1712 STTNLVDERQKTSGPSVASTGQAASAESLQ-HQTPSL-----ENLLARAMPHAFGRIAEN 1765

Qy        305 QEGPGSLAAGESSSLSAS----------TSVSDSSQKKEEHNYSLFVSDNLGEQPTKCSP 354
              ||       || |  :||          |:  |||:::: :  :   |: |   |  :|
Db       1766 QEPEDEPMGGEESDSAASMRSAASSNSQMSMGSSSQQQQQQDSDMTPRDSAG---TPSTP 1822

Qy        355 EEDE--------------------EDEEDVDDED---HDEGFGSEHELSENEEEEEEEE 390
               :|:                   | : | | ::    |:  |:  : |::|||| |
Db       1823 RDDKNQTLSVSAPDLAAARQRQASAETDGDADADETNSEDKTVGADDAMEEDDEEEETME 1882

Qy        391 DYEDDKDDDISDTFSEPGYENDSVEDLKEVTSISSRKRGKRRYFWEYSEQLTPSQQERML 450
              | ||| |||  ::      |:: | | |:        |:|  | :  | :|   | :
Db       1883 DEEDDDDDDDDES------SNENQEKLVELLG------GERGLFDKLKEVIT---GESLS 1927

Qy        451 RPSEWNRDTLPSNMYQKNGLHHGKYAVKKSRRTDV 485
                |  :|   :   :| |    |:  | |  |||
Db       1928 DASSSAKDATTNEAQKKGGKKPKKWFKKMSSYTDV 1962


RESULT 4
S48244
NMD2 protein - yeast (Saccharomyces cerevisiae)
N;Alternate names: IFS1 protein; protein YHR077c
C;Species: Saccharomyces cerevisiae
C;Date: 28-Oct-1994 #sequence_revision 10-Feb-1995 #text_change 09-Jul-2004
C;Accession: S48244; S46815; S64648; S64738
R;He, F.; Jacobson, A.
submitted to the EMBL Data Library, September 1994
A;Description: Identification of a novel component of the nonsense-mediated mRNA decay p
A;Reference number: S48244
A;Accession: S48244
A;Molecule type: DNA
A;Residues: 1-1089 <HEF>
A;Cross-references: UNIPROT:P38798; UNIPARC:UPI000013026A; EMBL:U14974; NID:g555938; PID
R;Favello, T.
submitted to the EMBL Data Library, June 1994
A;Description: The sequence of S. cerevisiae cosmid 9205.
A;Reference number: S46795
A;Accession: S46815
A;Molecule type: DNA
A;Residues: 'MYQQ',3-1089 <FAV>
A;Cross-references: UNIPARC:UPI0000168A23; EMBL:U10556; NID:g500825; PID:g500836; MIPS:Y
R;Lee, S.I.; Umen, J.G.; Varmus, H.E.
Proc. Natl. Acad. Sci. U.S.A. 92, 6587-6591, 1995
A;Title: A genetic screen identifies cellular factors involved in retroviral -1 frameshi
A;Reference number: S64648; MUID:95327692; PMID:7604038
A;Accession: S64648
A;Status: nucleic acid sequence not shown
A;Molecule type: DNA
A;Residues: 1-1089 <LEE>
A;Cross-references: UNIPARC:UPI000013026A; EMBL:U28158
R;Varmus, H.E.
submitted to the EMBL Data Library, May 1995
A;Reference number: S64738
A;Accession: S64738
A;Molecule type: DNA
A;Residues: 'MYQQ',3-1089 <VAR>
A;Cross-references: UNIPARC:UPI0000168A23; EMBL:U28158; NID:g967212; PID:g967213
C;Genetics:
A;Gene: SGD:NMD2; IFS1
A;Cross-references: SGD:S0001119; MIPS:YHR077c
A;Map position: 8R
A;Introns: 2/3

  Query Match             5.5%;  Score 184;  DB 2;  Length 1089;
  Best Local Similarity   20.2%;  Pred. No. 0.04;
  Matches   98;  Conservative   69;  Mismatches  159;  Indels  160;  Gaps   18;

Qy        133 ISKTPTLAQLNSEDSQSVSDSLYYPDSLFSVKQ--------NPLPSSFPGK-KITSRAAA 183
              |:  |    ||  |           ||: | ::         |  |::|  | |:  |
Db        707 INNQPNPFYLNYSDP---------PDNYFRIQLVTTILLNINRTPAAFTKKCKLLLR--- 754

Qy        184 PVCSSKTLQAEVPLSDCVQKASKPPSSTQIMVKT------NMYHNEKVNFHVECKDYVKK 237
                  |   | ||          |  |:  | :       |:: | |            ::
Db        755 -FFEYYTFIKEQPL----------PKETEFRVSSTFKKYENIFGNTK----------FER 793

Qy        238 AKVKINPVQQSRPLLSQIHTDAAKENTCYCGAVAKRQEKKGMEPLQGHATPALPFKETQE 297
              ::  :    :  ||  ::   :|::     : :    |:   |::             |  |
Db        794 SENLVESASRLESLLKSLNAIKSKDDRVKGSSASIHNGKESAVPIES-------ITEDDE 846

Qy        298 LLLSPLPQEGPGSLAAGESSSLSASTSVSDSSQKKEEHNYSLFVSDNLGEQPTKCSPEED 357
                  :   |    ||       ||  ::|:  | :              :  :   |:|
Db        847 ---DEDDENDDGVDLLGEDEDAEISTPNTESAPGKHQ-----------AKQDESEDEDD 891

Qy        358 EEDEEDVDDEDHDEGFGSEHELSENEEEEEEEEEDYEDDKDDDISDTFSEPGYENDSVEDL 417
              |:|:|| ||:|  |:     |    | :|:::|::|  ||| |::  |: |:  |  |  |
Db        892 EDDDEDDDDDDDDDDDDGE----EGDEDDDEDDDDDEDDDDEEEEDSDSDLEYGGDLDAD- 946

Qy        418 KEVTSISSRKRGKRRYFWEYSEQLTPSQQERMLRPSEWNRDTLPSNMYQKNGLHHGKYAV 477
                  |    :| : ||          || |:  |
Db        947 --------RDIEMKRMYEEY---------ERKLKDEE---------------------- 966

Qy        478 KKSRRTDVEDLTPNPKKLLQIGNELRKLNKVISDLTPVSELPLTAR----------PRSR 527
                  |   |:|    :|::|   : ||  ||::   ||   |:: :          | |
Db        967 ---ERKAEEELERQFQKMMQESIDARKSEKVVASKIPVISKPVSVQKPLLLKKSEEPSSS 1023

Qy        528 KE-------KNKLAF------------RACRL-------------KKKAQYEANKVKLWG 555
              ||          |:||            |  :|                ::| : | ||:|
Db       1024 KETYEELSKPKKIAFTFLTKSGKKTQSRILQLPTDVKFVSDVLEEEEKLKTERNKIKKIV 1083

Qy        556 LNTEYD 561
                |  :|
Db       1084 LKRSFD 1089


RESULT 5
DNCHNL
nucleolin - chicken
N;Alternate names: nucleolar protein C23
C;Species: Gallus gallus (chicken)
C;Date: 30-Sep-1991 #sequence_revision 30-Sep-1991 #text_change 09-Jul-2004
C;Accession: S08414; S10766; A32725; I50397; B30099
R;Maridor, G.; Nigg, E.A.
Nucleic Acids Res. 18, 1286, 1990
A;Title: cDNA sequences of chicken nucleolin/C23 and NO38/B23, two major nucleolar protei
A;Reference number: S08414; MUID:90206792; PMID:2320420
A;Accession: S08414
A;Status: translation not shown
A;Molecule type: mRNA
A;Residues: 1-694 <MAR>
A;Cross-references: UNIPROT:P15771; UNIPARC:UPI00001308C6; EMBL:X17199; NID:g63710; PIDN:
R;Maridor, G.; Krek, W.; Nigg, E.A.
```

```
Biochim. Biophys. Acta 1049, 126-133, 1990
A;Title: Structure and developmental expression of chicken nucleolin and NO38: coordinat
A;Reference number: S10766; MUID:90304215; PMID:2114180
A;Accession: S10766
A;Status: nucleic acid sequence not shown
A;Molecule type: mRNA
A;Residues: 1-694 <MAR2>
A;Cross-references: UNIPARC:UPI00001308C6
R;Peter, M.; Nakagawa, J.; Doree, M.; Labbe, J.C.; Nigg, E.A.
Cell 60, 791-801, 1990
A;Title: Identification of major nucleolar proteins as candidate mitotic substrates of c
A;Reference number: A32725; MUID:90182668; PMID:2178776
A;Accession: A32725
A;Molecule type: protein
A;Residues: 56-62;63-109;197-214 <PET>
A;Cross-references: UNIPARC:UPI00001739D2; UNIPARC:UPI00001739D3; UNIPARC:UPI00001739D4
R;Borer, R.A.; Lehner, C.F.; Eppenberger, H.M.; Nigg, E.A.
Cell 56, 379-390, 1989
A;Title: Major Nucleolar Proteins Shuttle between Nucleus and Cytoplasm.
A;Reference number: I50397; MUID:89119560; PMID:2914325
A;Accession: I50397
A;Status: translated from GB/EMBL/DDBJ
A;Molecule type: mRNA
A;Residues: 407-418,'R',420-519,'T',521-694 <BOR>
A;Cross-references: UNIPARC:UPI0000171376; GB:M21791; NID:g212411; PIDN:AAA48983.1; PID:
C;Comment: Phosphorylation of this protein by cdc2 kinase may contribute to the reorgani
C;Superfamily: nucleolin; ribonucleoprotein repeat homology
C;Keywords: DNA binding; duplication; nucleolus; nucleus; phosphoprotein; RNA binding; t
F;1-247/Domain: acidic <ACI>
F;54-91/Region: 7-residue repeats (T-P-A-K-K-A-[A/V])
F;254-262/Region: nuclear location signal
F;267-275/Region: nuclear location signal
F;282-347/Domain: ribonucleoprotein repeat homology <RRM1>
F;283-361/Domain: RNA binding #status predicted <RNA1>
F;372-435/Domain: ribonucleoprotein repeat homology <RRM2>
F;373-448/Domain: RNA binding #status predicted <RNA2>
F;462-525/Domain: ribonucleoprotein repeat homology <RRM3>
F;463-538/Domain: RNA binding #status predicted <RNA3>
F;554-618/Domain: ribonucleoprotein repeat homology <RRM4>
F;555-631/Domain: RNA binding #status predicted <RNA4>
F;632-694/Domain: glycine/arginine-rich <GRR>
F;56,63,70,77,85/Binding site: phosphate (Thr) (covalent) (by cdc2 kinase) #status predi

  Query Match             5.4%;  Score 180.5;  DB 1;  Length 694;
  Best Local Similarity   21.2%;  Pred. No. 0.035;
  Matches  117;  Conservative   74;  Mismatches  209;  Indels  151;  Gaps   23;

Qy        193 AEVPLSDCVQKASKPPSSTQIMVKTNMYHNEKVNFHVECKDYVKKA--KVKINPVQQSRP 250
              |: | :   ||   ||        |       |       | |  : :  :| : | :| :
Db          5 AKTPKNQMKQKKMAPPPK-----KVEESEEE------ESSDLEESSGEEVMVPPKKQQKA 53

Qy        251 LLSQIHTDAAKENTCYCGAVAKRQEKKGMEPLQGHATP-----ALPFKETQELLLSPLPQ 305
               :    | | |          |    ||   | :   ||     | | |:      ::| |:
Db         54 AV----TPAKK---------AATPAKKAATPAKKAVTPAKKAVATPAKKA----VAPSPK 96

Qy        306 E----GPGSLAAGESSSLSASTSVSDSSQKKEEHNYSLFVSDNLGEQPTKCSP------- 354
              :    | |:   |:::    |    :  :  ||       ||   |      |
Db         97 KAAVVGKGA-KNGKNAKKEESEEEDEDDEDEEDEDEEEESDEEEEPAVPVKPAAKKSAA 155

Qy        355 --------------------EEDEEDEEDVDDEDHDEGFG--------------------- 374
                                 |||:|:|:: |||  ||
Db        156 AVPAKKPAVVPAKQESEEEEEEDDEEEDEEDDESEDEAMDTTPAPVKKPTPAKATPAKAK 215

Qy        375 SEHELSENEEEEEEEEDYEDDKDDDISDTFSEPGYENDSVEDLKEVTSISSRKRGKRR-- 432
              :| |  |:||:|:|:|: |||:::|  ::  |   :    :  ||: : |: :  |::
Db        216 AESEDEEDEEDEDEDEEDEDDEEEDEEESEDEKPVKEAPGKRKKEMANKSAPEAKKKKTE 275

Qy        433 ----YFWEYSEQLTPSQQERMLRPSEWNRDTLPSNMYQKNGLHHGKYAVKKSRR------ 482
                  |  : : |||::    ||       |     : |  |   :  :  |:|
Db        276 TPASAFSLFVKNLTPTKDYEELR-------TAIKEFFGKKNLQVSEVRIGSSKRFGYVDF 328

Qy        483 TDVED----LTPNPKKLLQIGNELRKL--------NKVISDLTP--VSELPLTARPRSRK 528
                ||     |  | |||: :  :| |          ||   |      |  ||         |
Db        329 LSAEDMDKALQLNGKKLMGLEIKLEKAKSKESLKENKKERDARTLFVKNLPYRVTEDEMK 388

Qy        529 E--KNKLAFRACRLKK-----------KAQYEANKVKLWGLNTEYDNLLFVINSIKQEIV 575
                :| |  |    |:           | : || |        || |    ||:   ::
Db        389 NVPENALEVRLVLNKEGSSKGMAYIEFKTEAEAEKALEEKQGTEVDGRAMVIDYTGEKSQ 448

Qy        576 NRVQNPRDERGPNMGQKLEILIKDTLGLPVAGQTSEFVNQVLEKTA--------EGNPTG 627
                  |    ||       : : || : |    : :|   : :: :|          :| |
Db        449 QESQKGGGER------ESKTLIVNNLSYAASEET---LQELFKKATSIKMPQNNQGRPK- 498

Qy        628 GLVGLRIPTSK 638
                |   :  ||::
Db        499 GYAFVEFPTAE 509


RESULT 6
T38274
probable transcription initiation protein - fission yeast (Schizosaccharomyces pombe)
C;Species: Schizosaccharomyces pombe
C;Date: 03-Dec-1999 #sequence_revision 03-Dec-1999 #text_change 09-Jul-2004
C;Accession: T38274
R;Harris, D.; Squares, R.; Barrell, B.G.; Rajandream, M.A.; Wood, V.
submitted to the EMBL Data Library, September 1997
A;Reference number: Z21782
A;Accession: T38274
A;Status: preliminary; translated from GB/EMBL/DDBJ
A;Molecule type: DNA
A;Residues: 1-990 <HAR>
A;Cross-references: UNIPROT:O13936; UNIPARC:UPI0000061AF2; EMBL:Z99753; PIDN:CAB16890.1;
A;Experimental source: strain 972h-; cosmid c23C4
C;Genetics:
A;Gene: SPDB:SPAC23C4.19
A;Map position: 1
C;Keywords: transcription initiation

  Query Match             5.3%;  Score 179;  DB 2;  Length 990;
  Best Local Similarity   21.1%;  Pred. No. 0.064;
  Matches  107;  Conservative   72;  Mismatches  129;  Indels  198;  Gaps   23;

Qy        238 AKVKINPVQQS---RPLLSQIHTDAAKENTCYCGAVAKRQEKKGMEPLQGHATPALPFKE 294
              | ||||  : |      ||:   ::||| |           :|    |:  |
Db         26 ASVKINSTRASPNGSDLLND-DSEAAKITT----------NEKQSSPVDSH--------- 65

Qy        295 TQELLLSPLPQEGPG--SLAAGESSSLSASTSVSDSSQKKEEHNYSLFVSDNLGEQPTKC 352
                 | |   ::  ||  : :   :|::: :|::| |     | :|  |
Db         66 ----------NESPNDTTINKGEDGNENEVDNVNNNDKKEDEDN----VEENEEEADANE 111

Qy        353 SPEEDEEDEEDVDDEDHDEGFGSEHELSENE----------EEEEEEEDYEDDKDD-DIS 401
              |||||:|: |:|| ||  |   : : ::          | :|:||: ||::|:
Db        112 EEEEDEEDDEE-DEEDEDESGGGRRKRARHDRRNQFLDIEAEVDEDEEELEDEEDEIGRE 170

Qy        402 DTFSEPGYENDSVED----------LKEVTSISSRKRGKRRYFWEYSEQLTPSQ------ 445
              | | |      | | |           :|: |: : :  :     || |:   ||
Db        171 DGFIEEEVGADYVGDDRRHRELDRQRQELQSVDAERLAE-----EYREKYGRSQTVVGDT 225

Qy        446 ---QERMLRP-----------------------------------------SEWNRDTLP 461
                 :|:| |                                         | : ||:|
Db        226 SNVPQRLLLPSVNDPNIWAVRCKIGKEKDIVFTIMRKAMDLQYTSSPLEIISAFQRDSLV 285

Qy        462 SNMYQKNGLHHGKYAVKKSRRTD----VEDLTPNPKKLLQIGN--ELRKLNKVISDLTP- 514
                :| :        | |:|   |    | ::  |   |: |    :| |: | : :| |
Db        286 GYIYVE--------ARKQSHVLDALNGVLNVYTNNMILVPIKEMPDLLKVQKQVVELLPG 337

Qy        515 -------------------VSELPLTARPR---------SRKEKNKLAFRACRLKKKAQ- 545
                                :||  |||| |           | ||       ||  :::
Db        338 AYVRIRRGKYAGDLAQVDNLSENGLTARVRIVPRIDYSDGLKRKNSATRPQARLFNESEA 397

Qy        546 YEANKVKLWGLNTEYDNLLFVINSIKQEIVNRVQNPRDERGPNM----GQKLE--ILIKD 599
              :::|  |                        :||| :     ::  |   |:||
Db        398 FKSNPSKF---------------------------SKRGPRLFLFNNEEFEDGFLVKD 428

Qy        600 TLGLPVAGQTSEFVNQVLEKTAEGNP 625
                : ::   :| ||  |:: :: ||
```

```
Db        429 ---IRISSLITEGVNPTLDEVSKFNP 451


RESULT 7
A35804
nucleolin - human
N;Alternate names: phosphoprotein pp100; protein B50; protein C23
C;Species: Homo sapiens (man)
C;Date: 23-Oct-1990 #sequence_revision 23-Oct-1990 #text_change 09-Jul-2004
C;Accession: A35804; S04631; A48138; A55996
R;Srivastava, M.; McBride, O.W.; Fleming, P.J.; Pollard, H.B.; Burns, A.L.
J. Biol. Chem. 265, 14922-14931, 1990
A;Title: Genomic organization and chromosomal localization of the human nucleolin gene.
A;Reference number: A35804; MUID:90368666; PMID:2394707
A;Accession: A35804
A;Molecule type: DNA
A;Residues: 1-707 <SRI>
A;Cross-references: UNIPROT:P19338; UNIPARC:UPI0000161B7E; GB:M60858; GB:J05584; NID:g18
R;Srivastava, M.; Fleming, P.J.; Pollard, H.B.; Burns, A.L.
FEBS Lett. 250, 99-105, 1989
A;Title: Cloning and sequencing of the human nucleolin cDNA.
A;Reference number: S04631; MUID:89290043; PMID:2737305
A;Accession: S04631
A;Status: not compared with conceptual translation
A;Molecule type: mRNA
A;Residues: 1-707 <SR2>
A;Cross-references: UNIPARC:UPI0000161B7E
R;Ishikawa, F.; Matunis, M.J.; Dreyfuss, G.; Cech, T.R.
Mol. Cell. Biol. 13, 4301-4310, 1993
A;Title: Nuclear proteins that bind the pre-mRNA 3' splice site sequence r(UUAG/G) and t
A;Reference number: A48138; MUID:93309464; PMID:8321232
A;Accession: A48138
A;Molecule type: protein
A;Residues: 458-474 <ISH>
A;Cross-references: UNIPARC:UPI0000073927
A;Experimental source: HeLa cell nuclei
A;Note: sequence extracted from NCBI backbone (NCBIP:134645)
R;Jordan, P.; Heid, H.; Kinzel, V.; Kuebler, D.
Biochemistry 33, 14696-14706, 1994
A;Title: Major cell surface-located protein substrates of an ecto-protein kinase are hom
A;Reference number: A55996; MUID:95086063; PMID:7993898
A;Accession: A55996
A;Molecule type: protein
A;Residues: 231-236;349-362;399-403;458-461;655-656,'X',658-660 <JOR>
A;Cross-references: UNIPARC:UPI000017723C; UNIPARC:UPI000017723D; UNIPARC:UPI000017723E;
A;Experimental source: surface-labelled HeLa cells
C;Genetics:
A;Gene: GDB:NCL
A;Cross-references: GDB:125908; OMIM:164035
A;Map position: 2q12-2qter
C;Superfamily: nucleolin; ribonucleoprotein repeat homology
C;Keywords: DNA binding; nucleolus; nucleus; phosphoprotein; RNA binding
F;308-373/Domain: ribonucleoprotein repeat homology <RRM1>
F;394-456/Domain: ribonucleoprotein repeat homology <RRM2>
F;487-550/Domain: ribonucleoprotein repeat homology <RRM3>
F;573-634/Domain: ribonucleoprotein repeat homology <RRM4>

  Query Match             5.2%;  Score 174;  DB 2;  Length 707;
  Best Local Similarity   20.4%;  Pred. No. 0.078;
  Matches   99;  Conservative   76;  Mismatches  162;  Indels  148;  Gaps   20;

Qy        129 DEEVISKTPTLAQLNSEDSQSVSDSLYYPDSLFSVKQNPLPSSFPGKKITSRAAAPVCSS 188
              | : ::  |   : :||| :   |     ||  | ::  :|    ||| : :|  |  |
Db         13 DPKKMAPPPKEVEEDSEDEEMSEDE--EDDS--SGEEVVIPQK-KGKKAAATSAKKVVVS 67

Qy        189 KTLQAEVPLSDCVQKASKPPSSTQIMVKTNMYHNEKVNFHVECKDYVKKAKVKINPVQQS 248
              |   :| ::   :||:  |                          |     ||  : | :
Db         68 PT--KKVAVATPAKKAAVTPGK--------------------KAAATPAKKTVTPAK-- 102

Qy        249 RPLLSQIHTDAAKENTCYCGAVAKRQEKKGMEPLQGHATPALPFKETQELLLSPLPQEGP 308
                                  |||  |  | |  | | |:       : :| :|
Db        103 --------------------AVTTPGKKGATP--GKALVATPGKKG-----AAIPAKGA 134
```

```
Qy        309 GSLAAGESSSLSASTSVSDSSQKKEEHNYSLFVSDNLGEQP--------TKCSPEEDEED 360
              :   |:::    |    |   :::| :       |    :|             | :||:||
Db        135 KN---GKNAKKEDSDEEEDDDSEEDEEDDEDEDEDEDEIEPAAMKAAAAAPASEDEDDED 191

Qy        361 EEDVDDEDHDEGFGSEHE---------------------LSENEEEEEEEEDYEDDKDD 398
              :|| :|:| ||   || |                    ::|:|:|||::|| :|| |:
Db        192 DEDDEDDDDDEEDDSEEEAMETTPAKGKKAAKVVPVKAKNVAEDEDEEEDDEDEDDDDDE 251

Qy        399 DISDTFSEPGYENDSVEDLKEVTSISSRKRGKRRYFWEYSEQLTPSQQERMLRPSE---- 454
              |  |    |  | :  |: : |     ::: :       ::  |  ::: :  :|
Db        252 DDEDDDDEDDEEEEEEEEEEPVKEAPGKRKKEMA-----KQKAAPEAKKQKVEGTEPTTA 306

Qy        455 ---------WNRD-----TLPSNMYQKNGLHHGKYAVKKSRR------TDVEDLTPNPKK 494
                      :|:      |  |::: || |      :  :|:           |||    :|
Db        307 FNLFVGNLNFNKSAPELKTGISDVFAKNDLAVVDVRIGMTRKFGYVDFESAEDL----EK 362

Qy        495 LLQI------GNELRKLNKVISDLTPVSELPLTARPRSRKEKNKLAFRACRLKKKAQYEA 548
              |::      |||: || |                |: :  | :   |   | |   |:
Db        363 ALELTGLKVFGNEI-KLEK----------------PKGKDSKKERDARTL-LAKNLPYKV 404

Qy        549 NKVKL 553
              : :|
Db        405 TQDEL 409


RESULT 8
T46637
transcription factor 1, neural - rat
N;Alternate names: neural zinc finger factor-1
C;Species: Rattus norvegicus (Norway rat)
C;Date: 18-Feb-2000 #sequence_revision 18-Feb-2000 #text_change 09-Jul-2004
C;Accession: T46637
R;Jiang, Y.; Yu, V.C.; Buchholz, F.; O'Connell, S.; Rhodes, S.J.; Candeloro, C.; Xia, Y.F
J. Biol. Chem. 271, 10723-10730, 1996
A;Title: A novel family of Cys-Cys, His-Cys zinc finger transcription factors expressed i
A;Reference number: A58210; MUID:96210003; PMID:8631881
A;Accession: T46637
A;Status: preliminary; translated from GB/EMBL/DDBJ
A;Molecule type: mRNA
A;Residues: 1-1187 <JIA>
A;Cross-references: UNIPROT:P70475; UNIPARC:UPI00000E86A7; EMBL:U48809; NID:g1511631; PII
C;Genetics:
A;Gene: NZF-1
C;Function:
A;Description: binds specifically to a cis-regulatory element of the beta-retinoic acid r
e nervous system and in the pituitary gland
C;Keywords: DNA binding; transcription factor

  Query Match             5.0%;  Score 168.5;  DB 2;  Length 1187;
  Best Local Similarity   27.1%;  Pred. No. 0.28;
  Matches   73;  Conservative   45;  Mismatches  102;  Indels   49;  Gaps   11;

Qy        260 AKENTCYCGAVAKRQEKKGMEPLQGHATPALPFKETQELLLSPLPQEGPGSLAAGESSSL 319
              |:  : |   :||::: :  :|              ||    | |:  | :: | :|||:
Db         47 ARHRSVYGCPLAKKRKTQDKQP-------------QE----PAPKRKPFAVKA-DSSSV 87

Qy        320 SA--STSVSDSSQKKEEHNYSLFVSDNLGEQPTKCSPEEDEEDEEDV------DDEDHDE 371
                   :  ::    ||| :   |  ||  :       | | ||||::      |||| |:
Db         88 DECYESDGTEDMDDKEEDDDEEFSEDNDEQGDDDDEDEVDREDEEEIEEEDDEDDEDDDD 147

Qy        372 GFGSEHELSENEEEEEEEEDYEDDKDDDISDT--FSEPGYE---NDSVEDLKEVTSISSR 426
              |   | |  :::|||||||: |:::|  :| |    :|  :   ||  ::  |: : |
Db        148 GDDVEEEEDDDDEEEEEEEEEEENEDHQMSCTRIMQDPEKDDNNNDEYDNYDELVAKSLL 207

Qy        427 KRGKRRYFWEYSEQLTPSQQERMLRPSEWNRDTLPSNMYQ----KNGLHHGKYAVKKSRR 482
                ||          :     |    || | :|  ||  :     ||      |  :
Db        208 NLGK----------IAEDAAYRARTESEVNSNT--SNSLEDHSSKNENLGRKSELSLDLD 255

Qy        483 TDVEDLTPNPKKLLQIGNELRKLNKVISD 511
              :||   | :  |||  |: :  |:: |||
Db        256 SDVVRETVDSLKLLAQGHGV-VLSENISD 283
```

```
RESULT 9
A55320
immunophilin FKBP46 - fall armyworm
C;Species: Spodoptera frugiperda (fall armyworm)
C;Date: 06-Feb-1995 #sequence_revision 06-Feb-1995 #text_change 09-Jul-2004
C;Accession: A55320
R;Alnemri, E.S.; Fernandes-Alnemri, T.; Pomerenke, K.; Robertson, N.M.; Dudley, K.; DuBo
J. Biol. Chem. 269, 30828-30834, 1994
A;Title: FKBP46, a novel Sf9 insect cell nuclear immunophilin that forms a protein-kinas
A;Reference number: A55320; MUID:95074110; PMID:7527037
A;Accession: A55320
A;Status: preliminary
A;Molecule type: mRNA
A;Residues: 1-412 <ALN>
A;Cross-references: UNIPROT:Q26486; UNIPARC:UPI000012A81D; GB:U15038; NID:g595844; PID:g
C;Superfamily: yeast peptidylprolyl isomerase FPR3; BKBP-type peptidylprolyl isomerase h
C;Keywords: nucleus; phosphoprotein
F;324-371/Domain: BKBP-type peptidylprolyl isomerase homology <PPI>

  Query Match             5.0%;  Score 166.5;  DB 2;  Length 412;
  Best Local Similarity   20.6%;  Pred. No. 0.099;
  Matches   80;  Conservative   59;  Mismatches  118;  Indels  131;  Gaps   14;

Qy        274 QEKKGMEPLQGHATPALPFKETQEL----------LLSPLPQEGPGSLAAGESSSLSAST 323
              :|:   |  :  | | :| |  ::|             |  : |  |  | || |
Db        101 EEEAEEEEEEEEAPPLVPAKNKRKLENANDATANKKAKPDKKAGKNSAPAAESDS----- 155

Qy        324 SVSDSSQKKEEHNYSLFVSDNLGEQPTKCSPEEDEEDE-------EDVDDEDHDEGF--- 373
                                          ::|:||:       ||:| :::|| |
Db        156 ------------------------------DDDDEDQLQKFLDGEDIDTDENDESFKMN 184

Qy        374 -GSEHELSENEEEEEEEEDYEDDKDDDISDTFS------EPGYENDSVEDL-KEVTSISS 425
               :| : |: |:::|:||| ||| :||  :  :      :|  | ||  |  ||   :|
Db        185 TSAEGDDSDEEDDDEDEEDEEDDDEDDEEEEEAPKKKKKQPAAEQDSTLDTSKESVDMS- 243

Qy        426 RKRGKRRYFWEYSEQLTPSQQERMLRPSEWNRDTLPSNMYQKNGLHHGKYAVKKSRRTDV 485
                          :|: ||: |: :  :      |    | ||:   |   |:  :
Db        244 --------------KLSKSQKRRLKKKLQQQAKQQP----QVNGVDKPK---KEEPQQKA 282

Qy        486 EDLTPNPKKLLQIGNELRKLNKVISDLTPVSELPLTARPRSRKEKNKLAFRACRLKKKAQ 545
              |  |  ||  :     :|  | |:    : :| : : | ::  |  : :   |||:
Db        283 EKKKPEAKK--EEAPVEKKEKKQIAGGVSIEDLKVGSGPVAKAGKVVMVYYEGRLKQN-- 338

Qy        546 YEANKVKLWGLNTEYDNLLFVINSIKQEIVNRVQNPRDERGPNMGQKLEILIKDTL---- 601
                        |  :|| :                    :||  | |  :  |: :
Db        339 -----------NKMFDNCV-------------------KGP--GFKFRLGSKEVISGWD 365

Qy        602 ----GLPVAGQTSEFVNQVLEKTAEGNP 625
                  |: | |:        :   |:|:|
Db        366 VGIAGMKVGGKRKIVCPPAMAYGAKGSP 393


RESULT 10
A40253
acidic nuclear protein SPT5 - yeast (Saccharomyces cerevisiae)
N;Alternate names: protein YM9571.08; protein YML010w
C;Species: Saccharomyces cerevisiae
C;Date: 06-Dec-1991 #sequence_revision 06-Dec-1991 #text_change 09-Jul-2004
C;Accession: A40253; S55109
R;Swanson, M.S.; Malone, E.A.; Winston, F.
Mol. Cell. Biol. 11, 3009-3019, 1991
A;Title: SPT5, an essential gene important for normal transcription in Saccharomyces cer
A;Reference number: A40253; MUID:91246167; PMID:1840633
A;Accession: A40253
A;Molecule type: DNA
A;Residues: 1-1063 <SWA>
A;Cross-references: UNIPROT:P27692; UNIPARC:UPI0000053123; GB:M62882; NID:g172679; PIDN:
R;Gentles, S.; Bowman, S.
submitted to the EMBL Data Library, June 1995
A;Reference number: S55102
A;Accession: S55109
A;Molecule type: DNA
A;Residues: 1-1063 <GEN>
A;Cross-references: UNIPARC:UPI0000053123; EMBL:Z49810; NID:g854472; PIDN:CAA89942.1; PII
A;Experimental source: strain AB972
C;Genetics:
A;Gene: SGD:SPT5
A;Cross-references: SGD:S0004470; MIPS:YML010w
A;Map position: 13L
C;Keywords: nucleus; tandem repeat
F;148-216/Region: aspartic acid/glutamic acid-rich
F;931-1063/Region: 6-residue repeats

  Query Match             4.9%;  Score 165.5;  DB 2;  Length 1063;
  Best Local Similarity   23.2%;  Pred. No. 0.36;
  Matches   90;  Conservative   50;  Mismatches  109;  Indels  139;  Gaps   18;

Qy        246 QQSRPLLSQIHTDAAKENTCYCGAV----------AKRQEKKGMEPLQGHATPALPFKE 294
              | : |::    ||   |||     |           : :    : || |         |
Db         17 QFADPVVVPQSTDTKDENTSDKDTVDSGNVTTTESTERAESTSNIPPLDG---------E 67

Qy        295 TQELLLSPLPQEGPGSLAAGESSSLS--------ASTSVSDSSQ-----KKEEHNYSLFV 341
               :|    |   |     || |  | |          :|  :|||:     ||||
Db         68 EKEAKSEPQQPEDNAETAATEQVSSSNGPATDDAQATLNTDSSEANEIVKKEEG------ 121

Qy        342 SDNL---GEQPTKCS--------PEEDEEDEEDVDDEDHDEG-----------------F 373
              ||      |: || |          :||:|:|| ||:| ||
Db        122 SDERKRPREEDTKNSDGDTKDEGDNKDEDDDEDDDDDDDDEDDDDEAPTKRRRQERNRFL 181

Qy        374 GSEHELSENEEEEEEEEDYE----------DDKDDDISDTFSEPGYENDS------VEDL 417
                | |:|::|:|:|:||| |           ||:||:     | ||   |         :||
Db        182 DIEAEVSDDEDEDEDEEDSELVREGFITHGDDEDDEA----SAPGARRDDRLHRQLDQDL 237

Qy        418 KEVTSISSRKRGK---RRYFWEYSEQLTPSQQ-----ERMLRPS---------------- 453
               : :   :::  |    ||    |:|   : |      :| | ||
Db        238 NKTSEEDAQRLAKELRERYGRSSSKQYRAAAQDGYVPQRFLLPSVDTATIWGVRCRPGKE 297

Qy        454 ----------EWNRDTLPS-------NMYQKNGLHHGKYAVKKSRRTDVEDLTPNPKKLL 496
                       ::| |           :::|::  : |:  ::  ::: :|     |    :
Db        298 KELIRKLLKKKFNLDRAMGKKKLKILSIFQRDN-YTGRIYIEAPKQSVIEKFC-NGVPDI 355

Qy        497 QIGNELRKLNKVISDLTPVSELPLTARP 524
               |  :|          | || ||||  :|
Db        356 YISQKL---------LIPVQELPLLLKP 374


RESULT 11
A54514
glutamic acid-rich protein precursor - malaria parasite (Plasmodium falciparum)
N;Alternate names: GARP
C;Species: Plasmodium falciparum
C;Date: 15-Oct-1994 #sequence_revision 15-Oct-1994 #text_change 09-Jul-2004
C;Accession: A54514
R;Triglia, T.; Stahl, H.D.; Crewther, P.E.; Silva, A.; Anders, R.F.; Kemp, D.J.
Mol. Biochem. Parasitol. 31, 199-202, 1988
A;Title: Structure of a Plasmodium falciparum gene that encodes a glutamic acid-rich prot
A;Reference number: A54514; MUID:89040048; PMID:2903445
A;Accession: A54514
A;Status: preliminary
A;Molecule type: DNA
A;Residues: 1-678 <TRI>
A;Cross-references: UNIPROT:P13816; UNIPARC:UPI000012B0E0; GB:J03998; NID:g160298; PIDN:A
C;Genetics:
A;Introns: 25/3
C;Superfamily: histone H1
C;Keywords: tandem repeat

  Query Match             4.9%;  Score 165;  DB 2;  Length 678;
  Best Local Similarity   22.3%;  Pred. No. 0.22;
  Matches   61;  Conservative   44;  Mismatches  101;  Indels   68;  Gaps    9;

Qy        174 GKKITSRAAAPVCSSKTLQAEV-PLSDCVQKASKPPSSTQIMVKTNMYHNEKVNFHVE-- 230
              | :| :      |  : :  |  |||    |    |    :| |: :   |: |  ::
Db        454 GVEIINLEDKEACEEQHITVESRPLSQPQCKLIDEPEQLTLMDKSKV---EEKNLSIQEQ 510

Qy        231 ------------CKDYVKKAKVKINPVQQSRPLLSQIHTDAAKENTCYCGAVAKRQEKKG 278
```

```
                        :|  ||   ||   :    |  | | |  ::         |::| |
Db        511 LIGTIGRVNVVPRRDNHKKKMAKIEEAE----LQKQKHVDKEED---------KKEESKE 557

Qy        279 MEPLQGHATPALPFKETQELLLSPLPQEGPGSLAAGESSSLSASTSVSDSSQKKEEHNYS 338
              ::            || ||              :   :       :  :::||
Db        558 VQE---------ESKEVQE----------------DEEEVEEDEEEEEEEEEEEE-- 589

Qy        339 LFVSDNLGEQPTKCSPEEDEEDEEDVDDEDHDEGFGSEHELSENEEEEEEEEDYEDDKDD 398
                  |:  :   ||:||:||| |:|| |:    | :  |:|:: ||::| ||| ::
Db        590 --------EEEEEEEEEEEEEEEDEDEEDEDDAEEDEDDAEEDEDDAEEDDDEEDDDEE 641

Qy        399 DISDTFSEPGYENDSVEDLKEVTSISSRKRGKRR 432
              |  |   :   |::  |: :|  |    ||  |:
Db        642 D-DDEDEDEDEEDEEEEEEEEESEKKIKRNLRK 674


RESULT 12
S52522
hypothetical protein YPL009c - yeast (Saccharomyces cerevisiae)
N;Alternate names: hypothetical protein LPA8c; hypothetical protein YP8132.04c
C;Species: Saccharomyces cerevisiae
C;Date: 08-May-1995 #sequence_revision 21-Jul-1995 #text_change 09-Jul-2004
C;Accession: S52522; S59684
R;Badcock, K.; Churcher, C.
submitted to the EMBL Data Library, February 1995
A;Reference number: S52519
A;Accession: S52522
A;Molecule type: DNA
A;Residues: 1-1038 <BAD>
A;Cross-references: UNIPROT:Q12532; UNIPARC:UPI0000052E40; EMBL:Z48483; NID:g683777; PID
A;Experimental source: strain AB972
R;Hall, J.; Ahmed, A.; Bussey, H.; Fortin, N.; Friesen, J.D.; Storms, R.K.; Vo, D.H.; Wa
submitted to the EMBL Data Library, August 1995
A;Description: The sequence of Saccharomyces cerevisiae chromosome XVI left arm.
A;Reference number: S59677
A;Accession: S59684
A;Molecule type: DNA
A;Residues: 1-1038 <HAL>
A;Cross-references: UNIPARC:UPI0000052E40; EMBL:U33335; NID:g965076; PID:g965084; MIPS:Y
C;Genetics:
A;Cross-references: SGD:S0005930
A;Map position: 16L

  Query Match             4.9%;  Score 165;  DB 2;  Length 1038;
  Best Local Similarity   20.8%;  Pred. No. 0.37;
  Matches  103;  Conservative   73;  Mismatches  175;  Indels  144;  Gaps   21;

Qy        201 VQKASKPPSSTQIMVK----TNMYHNEKVNFHVECKDYVKKAKVKINPVQQSRPLL---S 253
              |      |  : ||  |     ::| :   | ||  |:  :| :|  | : |:  |     |
Db        588 VMMGKSPAETDQIYSKYIEDDDIYMSNSFNSHVWIKN-PEKTEVPPNTLMQAGILCMSSS 646

Qy        254 QIHTDAAKENTCYCGAVAKRQEKKGMEPLQGHATPALP---FKETQELLLSPLPQEGPGS 310
              :  :      :  :| |        | :     |     ||   |:  |   : ||   |
Db        647 EAWSKKISSSPWWCFA-------KNVSKFDGSDNSILPEGAFRLKNENDQNHLP---PAQ 696

Qy        311 LAAGESSSLSASTSVSDSSQKKEEHNYSLFVSDNLGEQPTKCSPEEDEEDEEDVDDEDHD 370
              |  |          || :: :   :|                ||:||:||: ::|: :
Db        697 LVMGFGFLWKVKTSGNEDNGDDDE-------------------EEEEEEEEEEEEEEE 736

Qy        371 EGFGSEHELSENEEEEEEEEDYEDDKDD----------DISDTFSEPGYENDSVE-DLKE 419
              |    | |  | ||||:|||  :|: |            | :|:     :|:|::| |:::
Db        737 E----EEEEEEKEEEEKEEEQQQDEDDSNEVNGLEKGGDSNDSTKNNSFEHDNLEKDIEK 792

Qy        420 ------------------------------------VTSISSRKRGKR----RYFWEYS 438
                                                 : :|:|  ||||    :   :|:
Db        793 HCTISSDTDSDSGNAKAKNDNSSTQRILDEPGVPISLIENINSNVRGKRGKLKKIQKKYA 852

Qy        439 EQLTPSQQERMLRPS--------EWNRDTLPSNMYQKNGLHHGKYAVKKSRRTDVEDLTP 490
              :|    : ||:||         |  :      : ::       |   :| ||      |
Db        853 DQ---DETERLLRLEALGTLKGIEKQQQRKKEEIMKREVREDRKNKREKQRRLQALKFTK 909

Qy        491 NPKKLLQIG---NELR-KLNK--VISDLTPVSELPLTARPRSRKEKNKLAFRACRLKKKA 544
                |  :      :||:  |:|  |: |: ||        |   | | |:  :    ||
```

```
Db        910 KEKARVNYDKHKSELKPSLDKGDVVDDIIPV----FAPWPALLKYKYKVKIQPGSAKKT- 964

Qy        545 QYEANKVKLWGLNTEYDNLLFVINSIKQEIV------NRVQNPRDE---RGPNMGQKLEI 595
                                     |  :::  |   :       | :  |::    :|      : :
Db        965 ----------------KTLTEILHYFKSRPLDGSSTDNEMDWPQEHEMIKGLKEQDLVLL 1008

Qy        596 LIKDTLGLPVAGQTS 610
              |  | | : :||| |
Db       1009 LCVDKLKVTIAGQKS 1023


RESULT 13
T50650
AP-3 complex beta3B chain [imported] - human
C;Species: Homo sapiens (man)
C;Date: 21-Jul-2000 #sequence_revision 21-Jul-2000 #text_change 09-Jul-2004
C;Accession: T50650
R;Newman, L.S.; McKeever, M.O.; Okano, H.J.; Darnell, R.B.
Cell 82, 773-783, 1995
A;Title: Beta-NAP, a cerebellar degeneration antigen, is a neuron-specific vesicle coat p
A;Reference number: Z25157; MUID:95401267; PMID:7671305
A;Accession: T50650
A;Status: preliminary; translated from GB/EMBL/DDBJ
A;Molecule type: mRNA
A;Residues: 1-1082 <NEW>
A;Cross-references: UNIPROT:Q13367; UNIPARC:UPI0000125030; EMBL:AF022152; PIDN:AAB71894.1

  Query Match             4.9%;  Score 164.5;  DB 2;  Length 1082;
  Best Local Similarity   22.1%;  Pred. No. 0.41;
  Matches   90;  Conservative   59;  Mismatches  150;  Indels  109;  Gaps   17;

Qy        176 KITSRAAAPVCSSKTLQ------AEVP--LSDCVQKASKPPSSTQIMVKTNMYHNEKVNF 227
              |:|      |:  :  |          ||     | ::| :|  :: : :||  :
Db        490 KLTDNIQVPMARASILWLIGEYCEHVPRIAPDVLRKMAKSPTAEEDIVKLQV-------I 542

Qy        228 HVECKDYVKKAKVKINPVQQSRPLLSQIHTDAAKENTCY------------------CGA 269
              ::  | |:  :|          ||:|     || :  |                    ||
Db        543 NLAAKLYLTNSK--------QTKLLTQYVLSLAKYDQNYDIRDRARFTRQLIVPSEQGGA 594

Qy        270 VAKRQEKKGMEPLQGHATPAL--PFKETQELLLSPL-----------------PQEGPGS 310
              :::  :|  : |       | |   ||:     |  |                 :| |
Db        595 LSRHAKKLFLAP---KPAPVLESSFKDRDHFQLGSLSHLLNAKATGYQELPDWPEEAPD- 650

Qy        311 LAAGESSSLSASTSVSDSSQKKEEHNYSLFVSDNLGEQPTKCSPEEDEEDEEDVDDEDHD 370
              :  :  :   |  |:  ::||:      | ||: ||      | : | | | : | :
Db        651 -PSVRNVEVPEWTKCSNREKRKEKE--KPFYSDSEGESGPTESADSDPESESESDSKSSS 707

Qy        371 E-GFGSEHELSENEEEEEEEE-----DYEDDKDDDISDT---FSEPGYENDSVEDLKEVT 421
              | | |     |:||:::|:||     :  |  ::|     |      |  | |  ::  : :
Db        708 ESGSGESSSESDNEDQDEDEEKGRGSESEQSEEDGKRKTKKKVPERKGEASSSDEGSDSS 767

Qy        422 SISSRKRGKRRYFWEYSEQLTPSQQERMLRPSEWNRDTLPSNMYQKNGLHHGKYAVKKSR 481
              | ||               ::|   :|  | |: |:| | ||:           | |:
Db        768 SSSS------------ESEMTSESEEEQLEPASWSRKTPPSS--------KSAPATKEIS 807

Qy        482 RTDVEDLTP------NPKKLLQI-------GNELRKLNKVISDLTPVS 516
                |:|| ||       :|  ::         |  |       | | |:|||
Db        808 LLDLEDFTPPSVQPVSPPAIVSTSLAADLEGLTLTDSTLVPSLLSPVS 855


RESULT 14
S22864
DNA topoisomerase (EC 5.99.1.2) I - Arabidopsis thaliana
C;Species: Arabidopsis thaliana (mouse-ear cress)
C;Date: 13-Jan-1995 #sequence_revision 13-Jan-1995 #text_change 09-Jul-2004
C;Accession: S22864
R;Kieber, J.K.; Signer, E.R.
submitted to the EMBL Data Library, September 1990
A;Description: Cloning and characterization of an Arabidopsis TOP1 gene.
A;Reference number: S22864
A;Accession: S22864
A;Molecule type: mRNA
A;Residues: 1-916 <KIE>
```

```
A;Cross-references: UNIPROT:P30181; UNIPARC:UPI000013712B; EMBL:X57544; NID:g16557; PID:
C;Genetics:
A;Gene: TOP1
C;Superfamily: eukaryotic type I DNA topoisomerase
C;Keywords: DNA binding; DNA replication; isomerase

  Query Match             4.9%;  Score 164;  DB 2;  Length 916;
  Best Local Similarity   21.1%;  Pred. No. 0.35;
  Matches  124;  Conservative   98;  Mismatches  226;  Indels  140;  Gaps   27;

Qy        111 GTKEVDYLGLDDFSSPYQDEEVIS--KTPTLAQLNSEDSQSVSD-SLYYPDSLFSVKQNP 167
              ||: |    :|: |    |:: ::  :  |:|  ||  |:| |  |   | : |  ::|
Db          2 GTETVSKPVMDNGSGDSDDDKPLAFKRNNTVAS-NSNQSKSNSQRSKAVPTTKVSPMRSP 60

Qy        168 LPS---SFPGKKIT-SRAAAPVCSSKTLQAEVPLSDCVQKASKPPSSTQIMVKTNMYHNE 223
              : |   : |  | :  ::: |  |||   |: || : :    |  |  |
Db         61 VTSPNGTTPSNKTSIVKSSMPSSSSKASPAKSPLRNDMPSTVKDRSQLQ---------- 109

Qy        224 KVNFHVECKDYVKKAKVKINPVQQSRPLLSQIHTDAAKENTCYCGAVAKRQEKKGMEPLQ 283
                        ||   : |::    :  ||| | :  :    ::    :    ::  |  ||
Db        110 --------KDQ-SECKIEHEDSEDDRPLSSILSGNKGPTSSRQVSSPQPEKKNNGDRPLD 160

Qy        284 GHATPALPFKETQELLLSPL-------PQEGPGSLAAGESSSLSASTSVSDSSQKKEEHN 336
              |:  :  :   |  :| :           |   |:  |   |      : |: | |  |
Db        161 -RASRIIKDESDDETPISSMFRKKIDSGMSGGNQLSNDEKKPLVQKLHQNGSTVKNEVPN 219

Qy        337 YSLFVSDNLGEQPTK---------------CSPEEDEEDEEDVDDEDHDEGFGSEHELS 380
               :    ||::| :                ||   :  :: |  |  |:|
Db        220 GKV-----LGKRPLEKNSSADQSSLKKAKISASPTSVKMKQDSVKKEIDDKGRVLVSPKM 274

Qy        381 ENEEEEEEEEDYEDDKDDD--ISDTF-SEPGYENDSVEDLKEV--TSISSRKRGKRRYFW 435
              : ::   |:  :|| |||  ||  | |:    | |    | |   | ||  : | |
Db        275 KAKQLSTREDGTDDDDDDDVPISKRFKSDSSNSNTSSAKPKAVKLNSTSSAAKPKARNVV 334

Qy        436 EYSEQLTPSQQERMLRPSEWNRD---------------TLPSN------MYQKNG---LH 471
               :      ::: : |:::                ||  |       |: :|   |:
Db        335 SPRSRAMTKNTKKVTKDSKYSTSSKSSPSSGDGQKKWTTLVHNGVIFPPPYKPHGIKILY 394

Qy        472 HGK--------------YAVKKSRRTDV-------EDLTPNPKKLLQIGNELRKLNKVIS 510
              ||               :||   | ||        |:   : ::||   : ::||:
Db        395 KGKPVDLTIEQEEVATMFAV--MRETDYYTKPQFRENFWNDWRRLLGKKHVIQKLDD--C 450

Qy        511 DLTPVSELPLTARPRSRK--EKNKLAFRACRLKKKAQYEANKVKLWGLNTEYDNLLFVIN 568
              | ||: |  |  : : ::   : | | :  ::|:: :|      :|          |::
Db        451 DFTPIYEWHLEEKEKKKQMSTEEKKALKEEKMKQEEKY------MWA----------VVD 494

Qy        569 SIKQEIVN-RVQNPRDERG----PNMGQ------KLEILIKDTLGLPV 605
              :|::| | ||: |   ||    | ||:         || :    | |:
Db        495 GVKEKIGNFRVEPPGLFRGRGEHPKMGKLKKRIHPCEITLNIGKGAPI 542


RESULT 15
S48529
NAB3 protein - yeast (Saccharomyces cerevisiae)
N;Alternate names: probable RNA/ssDNA-binding protein HMD1; protein P1945; protein YPL19
C;Species: Saccharomyces cerevisiae
C;Date: 10-Sep-1999 #sequence_revision 10-Sep-1999 #text_change 09-Jul-2004
C;Accession: S48529; S65209; S60122
R;Wilson, S.M.; Oberdorf, A.M.; Datar, K.V.; Swedlow, J.R.; Paddy, M.R.; Swanson, M.S.
submitted to the EMBL Data Library, January 1994
A;Description: Characterization of Nuclear Polyadenylated RNA-Binding Proteins from Sacc
A;Reference number: S48529
A;Accession: S48529
A;Molecule type: DNA
A;Residues: 1-802 <WIL>
A;Cross-references: UNIPROT:P38996; UNIPARC:UPI000004F979; EMBL:U05314; NID:g476219; PID
R;Rieger, M.; Mueller-Auer, S.; Schaefer, M.
submitted to the Protein Sequence Database, May 1996
A;Reference number: S65202
A;Accession: S65209
A;Molecule type: DNA
A;Residues: 1-802 <RIE>
A;Cross-references: UNIPARC:UPI000004F979; EMBL:Z73546; NID:g1370396; PID:g1370397; GSPD

A;Experimental source: strain S288C (AB972)
R;Sugimoto, K.; Matsumoto, K.; Kornberg, R.D.; Reed, S.I.; Wittenberg, C.
Mol. Gen. Genet. 248, 712-718, 1995
A;Title: Dosage suppressors of the dominant G1 cyclin mutant CLN3-2: identification of a
A;Reference number: S60122; MUID:96069710; PMID:7476874
A;Accession: S60122
A;Status: nucleic acid sequence not shown
A;Molecule type: DNA
A;Residues: 1-340,'I',342-802 <SUG>
A;Cross-references: UNIPARC:UPI00000694F8; GB:D37935; NID:g1235749; PID:g1235750
C;Genetics:
A;Gene: SGD:NAB3; NAB3; HMD1; MIPS:YPL190c
A;Cross-references: SGD:S0006111
A;Map position: 16L
C;Superfamily: yeast NAB3 protein; ribonucleoprotein repeat homology
C;Keywords: nucleus; RNA binding
F;331-396/Domain: ribonucleoprotein repeat homology <RRM1>

  Query Match             4.9%;  Score 163.5;  DB 1;  Length 802;
  Best Local Similarity   22.1%;  Pred. No. 0.32;
  Matches   76;  Conservative   57;  Mismatches  112;  Indels   99;  Gaps   14;

Qy        331 KKEEHNYSLFVSDNLGEQPTKCSPEEDEE-----DEEDVDDEDHDEGFGSEHELSENEEE 385
              ::|| |      ::: ::  :   |: ||     : |: ::|:| :  |:: :  :||||
Db         51 EREEENEEQHELEDVNDEEEEDKEEKGEENGEVINTEEEEEEHQQKGGNDDDDDNEEE 110

Qy        386 EEEEEDYEDDKDDDISD-----TFSEPGYENDSV-----------EDLKEVTSISSRKR 428
              |||||| :|| |||  |        | | :| ||           || |: |    ::
Db        111 EEEEEDDDDDDDDDDDEEEEEEEEEEGNDNSSVGSDSAAEDGEDEEDKKDKT--KDKEV 168

Qy        429 GKRRYFWEYSEQLTPSQQERMLRPSEWNRDTLPSNMYQKNGLHHGKYAVKKSRRTDVEDL 488
               ||   |  ||     :  : :   | :    |:||   |                 ||
Db        169 ELRRETLE-KEQKDVDEAIKKITREENDNTHFPTNMENVN----------------YDL 210

Qy        489 TPNPKKLLQIGNEL-----------RKLNKVISDLTPVSELPLTARP------------ 524
                | :   | |            |:: ::: |   |:  |:  |
Db        211 LQKQVKYIMDSNMLNLPQFQHLPQEEKMSAILAMLNSNSDTALSVPPHDSTISTTASASA 270

Qy        525 ----RSRKEK--------NKLAFRACRLKK---KAQYEANKVKLWGLN--TEYDNL---- 563
                 ||  ::         :: |    | |   : :::     | | |  ||  |:
Db        271 TSGARSNDQRKPPLSDAQRRMRFPRADLSKPITEEEHDRYAAYLHGENKITEMHNIPPKS 330

Qy        564 -LFV----INSIKQEIVNRVQNPRDERGPNMGQKLEILIKDTLG 602
               ||:    : :: :| : |: :|         |  ::| ||:  |
Db        331 RLPIGNLPLKNVSKEDLFRIFSP-------YGHIMQINIKNAFG 367


Search completed: February 28, 2006, 08:54:17
Job time : 41.9699 secs
```

GenCore version 5.1.7

OM protein - protein search, using sw model

Run on: February 28, 2006, 08:45:31 ; Search time 234.353 Seconds
(without alignments)
1923.730 Million cell updates/sec

Title: US-10-717-665A-44
Perfect score: 3349
Sequence: 1 MPQPSVSGMDPPFGDAFRSH.....TAEGNPTGGLVGLRIPTSKV 639

Scoring table: BLOSUM62
Gapop 10.0 , Gapext 0.5

Searched: 2166443 seqs, 705528306 residues

Total number of hits satisfying chosen parameters: 2166443

Minimum DB seq length: 0
Maximum DB seq length: 2000000000

Post-processing: Minimum Match 0%
Maximum Match 100%
Listing first 45 summaries

Database : UniProt_05.80:*
1: uniprot_sprot:*
2: uniprot_trembl:*

Pred. No. is the number of results predicted by chance to have a score greater than or equal to the score of the result being printed, and is derived by analysis of the total score distribution.

SUMMARIES

Result No.	Score	% Query Match	Length	DB	ID	Description
1	3328	99.4	639	2	Q8IUR6_HUMAN	Q8iur6 homo sapien
2	3326	99.3	639	2	Q8IZG1_HUMAN	Q8izg1 homo sapien
3	3136.5	93.7	640	2	Q8CDG5_MOUSE	Q8cdg5 mus musculu
4	3129	93.4	604	2	Q5HYK0_HUMAN	Q5hyk0 homo sapien
5	2148	64.1	408	2	Q5HYG4_HUMAN	Q5hyg4 homo sapien
6	2148	64.1	417	2	Q86YR3_HUMAN	Q86yr3 homo sapien
7	1704.5	50.9	351	2	Q8C8N7_MOUSE	Q8c8n7 mus musculu
8	1602.5	47.9	330	2	Q9CTQ7_MOUSE	Q9ctq7 mus musculu
9	1377.5	41.1	600	2	Q4RRX3_TETNG	Q4rrx3 tetraodon n
10	656	19.6	131	2	Q5ZMD5_CHICK	Q5zmd5 gallus gall
11	417	12.5	755	2	Q9VC61_DROME	Q9vc61 drosophila
12	386	11.5	293	2	Q7QG21_ANOGA	Q7qg21 anopheles g
13	227	6.8	2649	2	Q7RAS7_PLAYO	Q7ras7 plasmodium
14	212	6.3	761	2	Q7RBX2_PLAYO	Q7rbx2 plasmodium
15	210	6.3	707	2	Q8CE30_MOUSE	Q8ce30 mus musculu
16	209	6.2	706	1	NUCL_MOUSE	P09405 mus musculu
17	209	6.2	707	2	Q99K50_MOUSE	Q99k50 mus musculu
18	209	6.2	707	2	Q548M9_MOUSE	Q548m9 mus musculu
19	208	6.2	444	2	Q9CT46_MOUSE	Q9ct46 mus musculu
20	208	6.2	707	2	Q8CD23_MOUSE	Q8cd23 mus musculu
21	205	6.1	628	2	Q4Y1E8_PLACH	Q4y1e8 plasmodium
22	204.5	6.1	1852	2	Q9C2H4_NEUCR	Q9c2h4 neurospora
23	204	6.1	1182	2	Q8ID30_PLAF7	Q8id30 plasmodium
24	202	6.0	712	1	NUCL_RAT	P13383 rattus norv
25	201	6.0	676	2	Q8I510_PLAF7	Q8i510 plasmodium
26	199.5	6.0	734	2	Q8IDC4_PLAF7	Q8idc4 plasmodium
27	197.5	5.9	714	2	Q5U328_RAT	Q5u328 rattus norv
28	196.5	5.9	1162	2	Q98148_HHV8	Q98148 human herpe
29	195.5	5.8	771	2	Q54IN8_DICDI	Q54in8 dictyosteli
30	195	5.8	1708	2	Q8I413_PLAF7	Q8i413 plasmodium
31	195	5.8	2694	2	Q7RDP5_PLAYO	Q7rdp5 plasmodium
32	194.5	5.8	976	2	Q9DUN0_HHV8	Q9dun0 human herpe
33	194.5	5.8	1003	2	Q91LX9_HHV8	Q91lx9 human herpe
34	194	5.8	632	2	Q8I3T6_PLAF7	Q8i3t6 plasmodium
35	193.5	5.8	1156	2	Q8F2J8_LEPIN	Q8f2j8 leptospira
36	191.5	5.7	1129	2	Q9QR71_HHV8	Q9qr71 human herpe
37	191	5.7	1260	2	Q7RKZ2_PLAYO	Q7rkz2 plasmodium
38	189.5	5.7	727	1	YT521_HUMAN	Q96mu7 homo sapien
39	189	5.6	699	2	O57394_NARJA	O57394 narke japon
40	189	5.6	6473	2	Q8IKH9_PLAF7	Q8ikh9 plasmodium
41	188	5.6	2761	2	Q18447_CAEEL	Q18447 caenorhabdi
42	187.5	5.6	617	2	Q7RTH1_PLAYO	Q7rth1 plasmodium
43	187.5	5.6	1089	2	O40947_HHV8	O40947 human herpe
44	187	5.6	1922	2	Q54NP8_DICDI	Q54np8 dictyosteli
45	186.5	5.6	584	2	Q5D9Y2_SCHJA	Q5d9y2 schistosoma

ALIGNMENTS

RESULT 1
Q8IUR6_HUMAN
ID Q8IUR6_HUMAN PRELIMINARY; PRT; 639 AA.
AC Q8IUR6;
DT 01-MAR-2003 (TrEMBLrel. 23, Created)
DT 01-MAR-2003 (TrEMBLrel. 23, Last sequence update)
DT 01-MAR-2004 (TrEMBLrel. 26, Last annotation update)
DE Adult retina protein.
GN Name=LOC153222;
OS Homo sapiens (Human).
OC Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;
OC Mammalia; Eutheria; Euarchontoglires; Primates; Catarrhini; Hominidae;
OC Homo.
OX NCBI_TaxID=9606;
RN [1]
RP NUCLEOTIDE SEQUENCE.
RC TISSUE=Lung;
RX MEDLINE=22388257; PubMed=12477932; DOI=10.1073/pnas.242603899;
RA Strausberg R.L., Feingold E.A., Grouse L.H., Derge J.G.,
RA Klausner R.D., Collins F.S., Wagner L., Shenmen C.M., Schuler G.D.,
RA Altschul S.F., Zeeberg B., Buetow K.H., Schaefer C.F., Bhat N.K.,
RA Hopkins R.F., Jordan H., Moore T., Max S.I., Wang J., Hsieh F.,
RA Diatchenko L., Marusina K., Farmer A.A., Rubin G.M., Hong L.,
RA Stapleton M., Soares M.B., Bonaldo M.F., Casavant T.L., Scheetz T.E.,
RA Brownstein M.J., Usdin T.B., Toshiyuki S., Carninci P., Prange C.,
RA Raha S.S., Loquellano N.A., Peters G.J., Abramson R.D., Mullahy S.J.,
RA Bosak S.A., McEwan P.J., McKernan K.J., Malek J.A., Gunaratne P.H.,
RA Richards S., Worley K.C., Hale S., Garcia A.M., Gay L.J., Hulyk S.W.,
RA Villalon D.K., Muzny D.M., Sodergren E.J., Lu X., Gibbs R.A.,
RA Fahey J., Helton E., Ketteman M., Madan A., Rodrigues S., Sanchez A.,
RA Whiting M., Madan A., Young A.C., Shevchenko Y., Bouffard G.G.,
RA Blakesley R.W., Touchman J.W., Green E.D., Dickson M.C.,
RA Rodriguez A.C., Grimwood J., Schmutz J., Myers R.M.,
RA Butterfield Y.S.N., Krzywinski M.I., Skalska U., Smailus D.E.,
RA Schnerch A., Schein J.E., Jones S.J.M., Marra M.A.;
RT "Generation and initial analysis of more than 15,000 full-length human
RT and mouse cDNA sequences.";
RL Proc. Natl. Acad. Sci. U.S.A. 99:16899-16903(2002).
RN [2]
RP NUCLEOTIDE SEQUENCE.
RC TISSUE=Lung;
RA Director MGC Project;
RL Submitted (DEC-2002) to the EMBL/GenBank/DDBJ databases.
DR EMBL; BC041709; AAH41709.1; -; mRNA.
DR Ensembl; ENSG00000164463; Homo sapiens.
DR InterPro; IPR004827; TF_bZIP.
DR PROSITE; PS00036; BZIP_BASIC; UNKNOWN_1.
SQ SEQUENCE 639 AA; 72118 MW; ECFB92D9290DEDEB CRC64;

Query Match 99.4%; Score 3328; DB 2; Length 639;
Best Local Similarity 99.5%; Pred. No. 1.4e-148;
Matches 636; Conservative 0; Mismatches 3; Indels 0; Gaps 0;

Qy 1 MPQPSVSGMDPPFGDAFRSHTFSEQTLMSTDLLANSSDPDFMYELDREMNYQQNPRDNFL 60

```
Db          1 MPQPSVSGMDPPFGDAFRSHTFSEQTLMSTDLLANSSDPDFMYELDREMNYQQNPRDNFL 60

Qy         61 SLEDCKDIENLESFTDVLDNEGALTSNWEQWDTYCEDLTKYTKLTSCDIWGTKEVDYLGL 120

Db         61 SLEDCKDIENLESFTDVLDNEGALTSNWEQWDTYCEDLTKYTKLTSCDIWGTKEVDYLGL 120

Qy        121 DDFSSPYQDEEVISKTPTLAQLNSEDSQSVSDSLYYPDSLFSVKQNPLPSSFPGKKITSR 180

Db        121 DDFSSPYQDEEVISKTPTLAQLNSEDSQSVSDSLYYPDSLFSVKQNPLPSSFPGKKITSR 180

Qy        181 AAAPVCSSKTLQAEVPLSDCVQKASKPPSSTQIMVKTNMYHNEKVNFHVECKDYVKKAKV 240

Db        181 AAAPVCSSKTLQAEVPLSDCVQKASKPTSSTQIMVKTNMYHNEKVNFHVECKDYVKKAKV 240

Qy        241 KINPVQQSRPLLSQIHTDAAKENTCYCGAVAKRQEKKGMEPLQGHATPALPFKETQELLL 300

Db        241 KINPVQQSRPLLSQIHTDAAKENTCYCGAVAKRQEKKGMEPLQGHATPALPFKETQELLL 300

Qy        301 SPLPQEGPGSLAAGESSSLSASTSVSDSSQKKEEHNYSLFVSDNLGEQPTKCSPEEDEED 360

Db        301 SPLPQEGPGSLAAGESSSLSASTSVSDSSQKKEEHNYSLFVSDNLGEQPTKCSPEEDEED 360

Qy        361 EEDVDDEDHDEGFGSEHELSENEEEEEEEEDYEDDKDDDISDTFSEPGYENDSVEDLKEV 420

Db        361 EEDVDDEDHDEGFGSEHELSENEEEEEEEEDYEDDKDDDISDTFSEPGYENDSVEDLKEV 420

Qy        421 TSISSRKRGKRRYFWEYSEQLTPSQQERMLRPSEWNRDTLPSNMYQKNGLHHGKYAVKKS 480

Db        421 TSISSRKRGKRRYFWEYSEQLTPSQQERMLRPSEWNRDTLPSNMYQKNGLHHGKYAVKKS 480

Qy        481 RRTDVEDLTPNPKKLLQIGNELRKLNKVISDLTPVSELPLTARPRSRKEKNKLAFRACRL 540

Db        481 RRADVEDLTPNPKKLLQIGNELRKLNKVISDLTPVSELPLTARPRSRKEKNKLASRACRL 540

Qy        541 KKKAQYEANKVKLWGLNTEYDNLLFVINSIKQEIVNRVQNPRDERGPNMGQKLEILIKDT 600

Db        541 KKKAQYEANKVKLWGLNTEYDNLLFVINSIKQEIVNRVQNPRDERGPNMGQKLEILIKDT 600

Qy        601 LGLPVAGQTSEFVNQVLEKTAEGNPTGGLVGLRIPTSKV 639

Db        601 LGLPVAGQTSEFVNQVLEKTAEGNPTGGLVGLRIPTSKV 639
```

RESULT 2
Q8IZG1_HUMAN
ID Q8IZG1_HUMAN PRELIMINARY; PRT; 639 AA.
AC Q8IZG1;
DT 01-MAR-2003 (TrEMBLrel. 23, Created)
DT 01-MAR-2003 (TrEMBLrel. 23, Last sequence update)
DT 01-MAR-2004 (TrEMBLrel. 26, Last annotation update)
DE Adult retina protein.
OS Homo sapiens (Human).
OC Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;
OC Mammalia; Eutheria; Euarchontoglires; Primates; Catarrhini; Hominidae;
OC Homo.
OX NCBI_TaxID=9606;
RN [1]
RP NUCLEOTIDE SEQUENCE.
RA Li F., Yao K.T.;
RL Submitted (AUG-2002) to the EMBL/GenBank/DDBJ databases.
DR EMBL; AY139008; AAN28956.1; -; mRNA.
DR InterPro; IPR004827; TF_bZIP.
DR PROSITE; PS00036; BZIP_BASIC; UNKNOWN_1.
SQ SEQUENCE 639 AA; 72090 MW; 3A19E0926B9A6406 CRC64;

Query Match 99.3%; Score 3326; DB 2; Length 639;
Best Local Similarity 99.5%; Pred. No. 1.7e-148;
Matches 636; Conservative 0; Mismatches 3; Indels 0; Gaps 0;

```
Qy          1 MPQPSVSGMDPPFGDAFRSHTFSEQTLMSTDLLANSSDPDFMYELDREMNYQQNPRDNFL 60

Db          1 MPQPSVSGMDPPFGDAFRSHTFSEQTLMSTDLLANSSDPDFMYELDREMNYQQNPRDNFL 60

Qy         61 SLEDCKDIENLESFTDVLDNEGALTSNWEQWDTYCEDLTKYTKLTSCDIWGTKEVDYLGL 120

Db         61 SLEDCKDIENLESFTDVLDNEGALTSNWEQWDTYCEDLTKYTKLTSCDIWGTKEVDYLGL 120

Qy        121 DDFSSPYQDEEVISKTPTLAQLNSEDSQSVSDSLYYPDSLFSVKQNPLPSSFPGKKITSR 180

Db        121 DDFSSPYQDEEVISKTPTLAQLNSEDSQSVSDSLYYPDSLFSVKQNPLPSSFPGKKITSR 180

Qy        181 AAAPVCSSKTLQAEVPLSDCVQKASKPPSSTQIMVKTNMYHNEKVNFHVECKDYVKKAKV 240

Db        181 AAAPVCSSKTLQAEVPLSDCVQKASKPTSSTQIMVKTNMYHNEKVNFHVECKDYVKKAKV 240

Qy        241 KINPVQQSRPLLSQIHTDAAKENTCYCGAVAKRQEKKGMEPLQGHATPALPFKETQELLL 300

Db        241 KINPVQQSRPLLSQIHTDAAKENTCYCGAVAKRQEKKGMEPLQGHATPALPFKETQELLL 300

Qy        301 SPLPQEGPGSLAAGESSSLSASTSVSDSSQKKEEHNYSLFVSDNLGEQPTKCSPEEDEED 360

Db        301 SPLPQEGPGSLAAGESSSLSASTSVSDSSQKKEEHNYSLFVSDNLGEQPTKCSPEEDEED 360

Qy        361 EEDVDDEDHDEGFGSEHELSENEEEEEEEEDYEDDKDDDISDTFSEPGYENDSVEDLKEV 420

Db        361 EEDVDDEDHDEGFGSEHELSENEEEEEEEEDYEDDKDDDISDTFSEPGYENDSVEDLKEV 420

Qy        421 TSISSRKRGKRRYFWEYSEQLTPSQQERMLRPSEWNRDTLPSNMYQKNGLHHGKYAVKKS 480

Db        421 TSISSRKRGKRRYFWEYSEQLTPSQQERMLRPSEWNRDTLPSNMYQKNGLHHGKYAVKKS 480

Qy        481 RRTDVEDLTPNPKKLLQIGNELRKLNKVISDLTPVSELPLTARPRSRKEKNKLAFRACRL 540

Db        481 RRTDVEDLTPNPKKLLQIGNELRKLNKVISDLTPVSELPLTARPRSRKEKNKLASRACRL 540

Qy        541 KKKAQYEANKVKLWGLNTEYDNLLFVINSIKQEIVNRVQNPRDERGPNMGQKLEILIKDT 600

Db        541 KKKAQYEANKVKLWGLNTEYGNLLFVINSIKQEIVNRVQNPRDERGPNMGQKLEILIKDT 600

Qy        601 LGLPVAGQTSEFVNQVLEKTAEGNPTGGLVGLRIPTSKV 639

Db        601 LGLPVAGQTSEFVNQVLEKTAEGNPTGGLVGLRIPTSKV 639
```

RESULT 3
Q8CDG5_MOUSE
ID Q8CDG5_MOUSE PRELIMINARY; PRT; 640 AA.
AC Q8CDG5;
DT 01-MAR-2003 (TrEMBLrel. 23, Created)
DT 01-MAR-2003 (TrEMBLrel. 23, Last sequence update)
DT 01-MAR-2004 (TrEMBLrel. 26, Last annotation update)
DE Mus musculus adult male testis cDNA, RIKEN full-length enriched
DE library, clone:4932441F15 product:hypothetical protein, full insert
DE sequence.
GN Name=A930001N09Rik;
OS Mus musculus (Mouse).
OC Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;
OC Mammalia; Eutheria; Euarchontoglires; Glires; Rodentia; Sciurognathi;
OC Muroidea; Muridae; Murinae; Mus.
OX NCBI_TaxID=10090;
RN [1]
RP NUCLEOTIDE SEQUENCE.
RC STRAIN=C57BL/6J; TISSUE=Testis;
RX MEDLINE=99279253; PubMed=10349636; DOI=10.1016/S0076-6879(99)03004-9;
RA Carninci P., Hayashizaki Y.;
RT "High-efficiency full-length cDNA cloning.";
RL Meth. Enzymol. 303:19-44(1999).
RN [2]
RP NUCLEOTIDE SEQUENCE.
RC STRAIN=C57BL/6J; TISSUE=Testis;
RX MEDLINE=21085660; PubMed=11217851; DOI=10.1038/35055500;
RA Kawai J., Shinagawa A., Shibata K., Yoshino M., Itoh M., Ishii Y.,
RA Arakawa T., Hara A., Fukunishi Y., Konno H., Adachi J., Fukuda S.,
RA Aizawa K., Izawa M., Nishi K., Kiyosawa H., Kondo S., Yamanaka I.,
RA Saito T., Okazaki Y., Gojobori T., Bono H., Kasukawa T., Saito R.,

```
RA   Kadota K., Matsuda H.A., Ashburner M., Batalov S., Casavant T.,
RA   Fleischmann W., Gaasterland T., Gissi C., King B., Kochiwa H.,
RA   Kuehl P., Lewis S., Matsuo Y., Nikaido I., Pesole G., Quackenbush J.,
RA   Schriml L.M., Staubli F., Suzuki R., Tomita M., Wagner L., Washio T.,
RA   Sakai K., Okido T., Furuno M., Aono H., Baldarelli R., Barsh G.,
RA   Blake J., Boffelli D., Bojunga N., Carninci P., de Bonaldo M.F.,
RA   Brownstein M.J., Bult C., Fletcher C., Fujita M., Gariboldi M.,
RA   Gustincich S., Hill D., Hofmann M., Hume D.A., Kamiya M., Lee N.H.,
RA   Lyons P., Marchionni L., Mashima J., Mazzarelli J., Mombaerts P.,
RA   Nordone P., Ring B., Ringwald M., Rodriguez I., Sakamoto N.,
RA   Sasaki H., Sato K., Schoenbach C., Seya T., Shibata Y., Storch K.-F.,
RA   Suzuki H., Toyo-oka K., Wang K.H., Weitz C., Whittaker C., Wilming L.,
RA   Wynshaw-Boris A., Yoshida K., Hasegawa Y., Kawaji H., Kohtsuki S.,
RA   Hayashizaki Y.;
RT   "Functional annotation of a full-length mouse cDNA collection.";
RL   Nature 409:685-690(2001).
RN   [3]
RP   NUCLEOTIDE SEQUENCE.
RC   STRAIN=C57BL/6J; TISSUE=Testis;
RA   The FANTOM Consortium,
RA   the RIKEN Genome Exploration Research Group Phase I & II Team;
RT   "Analysis of the mouse transcriptome based on functional annotation of
RT   60,770 full-length cDNAs.";
RL   Nature 420:563-573(2002).
RN   [4]
RP   NUCLEOTIDE SEQUENCE.
RC   STRAIN=C57BL/6J; TISSUE=Testis;
RX   MEDLINE=20499374; PubMed=11042159; DOI=10.1101/gr.145100;
RA   Carninci P., Shibata Y., Hayatsu N., Sugahara Y., Shibata K., Itoh M.,
RA   Konno H., Okazaki Y., Muramatsu M., Hayashizaki Y.;
RT   "Normalization and subtraction of cap-trapper-selected cDNAs to
RT   prepare full-length cDNA libraries for rapid discovery of new genes.";
RL   Genome Res. 10:1617-1630(2000).
RN   [5]
RP   NUCLEOTIDE SEQUENCE.
RC   STRAIN=C57BL/6J; TISSUE=Testis;
RX   MEDLINE=20530913; PubMed=11076861; DOI=10.1101/gr.152600;
RA   Shibata K., Itoh M., Aizawa K., Nagaoka S., Sasaki N., Carninci P.,
RA   Konno H., Akiyama J., Nishi K., Kitsunai T., Tashiro H., Itoh M.,
RA   Sumi N., Ishii Y., Nakamura S., Hazama M., Nishine T., Harada A.,
RA   Yamamoto R., Matsumoto H., Sakaguchi S., Ikegami T., Kashiwagi K.,
RA   Fujiwake S., Inoue K., Togawa Y., Izawa M., Ohara E., Watahiki M.,
RA   Yoneda Y., Ishikawa T., Ozawa K., Tanaka T., Matsuura S., Kawai J.,
RA   Okazaki Y., Muramatsu M., Inoue Y., Kira A., Hayashizaki Y.;
RT   "RIKEN integrated sequence analysis (RISA) system-384-format
RT   sequencing pipeline with 384 multicapillary sequencer.";
RL   Genome Res. 10:1757-1771(2000).
RN   [6]
RP   NUCLEOTIDE SEQUENCE.
RC   STRAIN=C57BL/6J; TISSUE=Testis;
RA   Adachi J., Aizawa K., Akimura T., Arakawa T., Bono H., Carninci P.,
RA   Fukuda S., Furuno M., Hanagaki T., Hara A., Hashizume W.,
RA   Hayashida K., Hayatsu N., Hiramoto K., Hiraoka T., Hirozane T.,
RA   Hori F., Imotani K., Ishii Y., Itoh M., Kagawa I., Kasukawa T.,
RA   Katoh H., Kawai J., Kojima Y., Kondo S., Konno H., Kouda M., Koya S.,
RA   Kurihara C., Matsuyama T., Miyazaki A., Murata M., Nakamura M.,
RA   Nishi K., Nomura K., Numazaki R., Ohno M., Ohsato N., Okazaki Y.,
RA   Saito R., Saitoh H., Sakai C., Sakai K., Sakazume N., Sano H.,
RA   Sasaki D., Shibata K., Shinagawa A., Shiraki T., Sogabe Y., Tagami M.,
RA   Tagawa A., Takahashi F., Takaku-Akahira S., Takeda Y., Tanaka T.,
RA   Tomaru A., Toya T., Yasunishi A., Muramatsu M., Hayashizaki Y.;
RL   Submitted (JUL-2001) to the EMBL/GenBank/DDBJ databases.
DR   EMBL; AK030092; BAC26779.1; -; mRNA.
DR   Ensembl; ENSMUSG00000048249; Mus musculus.
DR   MGI; MGI:1924378; A930001N09Rik.
DR   InterPro; IPR004827; TF_bZIP.
DR   PROSITE; PS00036; BZIP_BASIC; UNKNOWN_1.
KW   Hypothetical protein.
SQ   SEQUENCE   640 AA;  72598 MW;  FE02C532FA34E1DE CRC64;

  Query Match           93.7%;  Score 3136.5;  DB 2;  Length 640;
  Best Local Similarity 93.4%;  Pred. No. 1.4e-139;
  Matches  598;  Conservative   19;  Mismatches   22;  Indels    1;  Gaps    1;

Qy          1 MPQPSVSGMDPPFGDAFRSHTFSEQTLMSTDLLANSSDPDFMYELDREMNYQQNPRDNFL 60
              ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db          1 MPQPSVSGMDPPFGDAFRSHTFSEQTLMSTDLLANSSDPDFMYELDREMNYQQNPRDNFL 60

Qy         61 SLEDCKDIENLESFTDVLDNEGALTSNWEQWDTYCEDLTKYTKLTSCDIWGTKEVDYLGL 120
              |||||||||||:|||||||| ||||||||||||||||||||||||||||||||||||||||
Db         61 SLEDCKDIENLETFTDVLDNEDALTSNWEQWDTYCEDLTKYTKLTSCDIWGTKEVDYLGL 120

Qy        121 DDFSSPYQDEEVISKTPTLAQLNSEDSQSVSDSLYYPDSLFSVKQNPL-PSSFPGKKITS 179
              |||||||||||||||||||||||||||||||||||||||||||||||| ||||| ||||:
Db        121 DDFSSPYQDEEVISKTPTLAQLNSEDSQSVSDSLYYPDSLFSVKQNPLPPSSFPSKKITN 180

Qy        180 RAAAPVCSSKTLQAEVPLSDCVQKASKPPSSTQIMVKTNMYHNEKVNFHVECKDYVKKAK 239
              |||||||||||||||| |||||||||| ||||||||||||||||||||||||||||||||
Db        181 RAAAPVCSSKTLQAEVPSSDCVQKASKPTSSTQIMVKTNMYHNEKVNFHVECKDYVKKAK 240

Qy        240 VKINPVQQSRPLLSQIHTDAAKENTCYCGAVAKRQEKKGMEPLQGHATPALPFKETQELL 299
              |||||||| |||||:| ||||||||||||||||::|:|| ||  |||||||||||||||
Db        241 VKINPVQQGRPLLSQVHIDAAKENTCYCGAVAKRQERRGVEPHQGRGTPALPFKETQELL 300

Qy        300 LSPLPQEGPGSLAAGESSSLSASTSVSDSSQKKEEHNYSLFVSDNLGEQPTKCSPEEDEE 359
              |||| |: || :|  || |||||||||||||||||||||||||: ||||| |||:||:
Db        301 LSPLTQDSPGLVATAESGSLSASTSVSDSSQKKEEHNYSLFVSDNMREQPTKYSPEDDED 360

Qy        360 DEEDVDDEDHDEGFGSEHELSENEEEEEEEEDYEDDKDDDISDTFSEPGYENDSVEDLKE 419
              ||:: |||||||||||||||||||||||||||||:|||||||||||||||||||||||||
Db        361 DEDEFDDEDHDEGFGSEHELSENEEEEEEEEDYEDDRDDDISDTFSEPGYENDSVEDLKE 420

Qy        420 VTSISSRKRGKRRYFWEYSEQLTPSQQERMLRPSEWNRDTLPSNMYQKNGLHHGKYAVKK 479
              :|||||||||||||||||||||||||||||:|||||||||||||||||||||||||||||
Db        421 MTSISSRKRGKRRYFWEYSEQLTPSQQERILRPSEWNRDTLPSNMYQKNGLHHGKYAVKK 480

Qy        480 SRRTDVEDLTPNPKKLLQIGNELRKLNKVISDLTPVSELPLTARPRSRKEKNKLAFRACR 539
              |||||||||||||||||||||||||||||||||||||||||||||||||||||||| ||||
Db        481 SRRTDVEDLTPNPKKLLQIGNELRKLNKVISDLTPVSELPLTARPRSRKEKNKLASRACR 540

Qy        540 LKKKAQYEANKVKLWGLNTEYDNLLFVINSIKQEIVNRVQNPRDERGPNMGQKLEILIKD 599
              |||||||||||||||||||||||||||||||||:|||||||:|| |:||||||||||||
Db        541 LKKKAQYEANKVKLWGLNTEYDNLLFVINSIKQDIVNRVQNPREEREPSMGQKLEILIKD 600

Qy        600 TLGLPVAGQTSEFVNQVLEKTAEGNPTGGLVGLRIPTSKV 639
              |||||||||||||||| |||||||||||||||| |||
Db        601 TLGLPVAGQTSEFVNQVLGKTAEGNPTGGLVGLRIPASKV 640


RESULT 4
Q5HYK0_HUMAN
ID   Q5HYK0_HUMAN PRELIMINARY;      PRT;   604 AA.
AC   Q5HYK0;
DT   10-MAY-2005 (TrEMBLrel. 30, Created)
DT   10-MAY-2005 (TrEMBLrel. 30, Last sequence update)
DT   10-MAY-2005 (TrEMBLrel. 30, Last annotation update)
DE   Hypothetical protein DKFZp313F2319 (Fragment).
GN   Name=DKFZp313F2319;
OS   Homo sapiens (Human).
OC   Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;
OC   Mammalia; Eutheria; Euarchontoglires; Primates; Catarrhini; Hominidae;
OC   Homo.
OX   NCBI_TaxID=9606;
RN   [1]
RP   NUCLEOTIDE SEQUENCE.
RC   TISSUE=Adipose;
RG   The German cDNA Consortium;
RA   Koehrer K., Beyer A., Mewes H.W., Weil B., Amid C., Osanger A.,
RA   Fobo G., Han M., Wiemann S.;
RL   Submitted (JAN-2005) to the EMBL/GenBank/DDBJ databases.
DR   EMBL; BX647573; CAI46104.1; -; mRNA.
DR   InterPro; IPR004827; TF_bZIP.
DR   PROSITE; PS00036; BZIP_BASIC; UNKNOWN_1.
KW   Hypothetical protein.
FT   NON_TER     604    604
SQ   SEQUENCE   604 AA;  68572 MW;  977C229B63E2E4C2 CRC64;
```

```
  Query Match           93.4%;  Score 3129;  DB 2;  Length 604;
  Best Local Similarity   99.0%;  Pred. No. 2.9e-139;
  Matches  598;  Conservative    2;  Mismatches    2;  Indels    2;  Gaps    1;

Qy          1 MPQPSVSGMDPPFGDAFRSHTFSEQTLMSTDLLANSSDPDFMYELDREMNYQQNPRDNFL 60
              ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db          1 MPQPSVSGMDPPFGDAFRSHTFSEQTLMSTDLLANSSDPDFMYELDREMNYQQNPRDNFL 60

Qy         61 SLEDCKDIENLESFTDVLDNEGALTSNWEQWDTYCEDLTKYTKLTSCDIWGTKEVDYLGL 120
              ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db         61 SLEDCKDIENLESFTDVLDNEGALTSNWEQWDTYCEDLTKYTKLTSCDIWGTKEVDYLGL 120

Qy        121 DDFSSPYQDEEVISKTPTLAQLNSEDSQSVSDSLYYPDSLFSVKQNPLPSSFPGKKITSR 180
              ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db        121 DDFSSPYQDEEVISKTPTLAQLNSEDSQSVSDSLYYPDSLFSVKQNPLPSSFPGKKITSR 180

Qy        181 AAAPVCSSKTLQAEVPLSDCVQKASKPPSSTQIMVKTNMYHNEKVNFHVECKDYVKKAKV 240
              |||||||||||||||||||||||| |||||||||||||||||:||||||||
Db        181 AAAPVCSSKTLQAEVPLSDCVQKASKPTSSTQIMVKTNMYHNEKVNFHVECRDYVKKAKV 240

Qy        241 KINPVQQSRPLLSQIHTDAAKENTCYCGAVAKRQEKKGMEPLQGHATPALPFKETQELLL 300
              ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db        241 KINPVQQSRPLLSQIHTDAAKENTCYCGAVAKRQEKKGMEPLQGHATPALPFKETQELLL 300

Qy        301 SPLPQEGPGSLAAGESSSLSASTSVSDSSQKKEEHNYSLFVSDNLGEQPTKCSPEEDEED 360
              ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db        301 SPLPQEGPGSLAAGESSSLSASTSVSDSSQKKEEHNYSLFVSDNLGEQPTKCSPEEDEED 360

Qy        361 EEDVDDEDHDEGFGSEHELSENEEEEEEEEDYEDDKDDDISDTFSEPGYENDSVEDLKEV 420
              ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db        361 EEDVDDEDHDEGFGSEHELSENEEEEEEEEDYEDDKDDDISDTFSEPGYENDSVEDLKEV 420

Qy        421 TSISSRKRGKRRYFWEYSEQLTPSQQERMLRPSEWNRDTLPSNMYQKNGLHH--GKYAVK 478
              |||||||||||||||||||||||||||||||||||||||||||||||||  ||||||
Db        421 TSISSRKRGKRRYFWEYSEQLTPSQQERMLRPSEWNRDTLPSNMYQKNGLHHGKGKYAVK 480

Qy        479 KSRRTDVEDLTPNPKKLLQIGNELRKLNKVISDLTPVSELPLTARPRSRKEKNKLAFRAC 538
              |||||||||||||||||||||||||||||||||||||||||||||||||||||||||| |||
Db        481 KSRRTDVEDLTPNPKKLLQIGNELRKLNKVISDLTPVSELPLTARPRSRKEKNKLASRAC 540

Qy        539 RLKKKAQYEANKVKLWGLNTEYDNLLFVINSIKQEIVNRVQNPRDERGPNMGQKLEILIK 598
              ||||||||||||||||||||||||||||||||||||||||||||||||:||||||||||
Db        541 RLKKKAQYEANKVKLWGLNTEYDNLLFVINSIKQEIVNRVQNPRDERGPDMGQKLEILIK 600

Qy        599 DTLG 602
              ||||
Db        601 DTLG 604
```

```
RESULT 5
Q5HYG4_HUMAN
ID   Q5HYG4_HUMAN PRELIMINARY;      PRT;   408 AA.
AC   Q5HYG4;
DT   10-MAY-2005 (TrEMBLrel. 30, Created)
DT   10-MAY-2005 (TrEMBLrel. 30, Last sequence update)
DT   10-MAY-2005 (TrEMBLrel. 30, Last annotation update)
DE   Hypothetical protein DKFZp686G2059 (Fragment).
GN   Name=DKFZp686G2059;
OS   Homo sapiens (Human).
OC   Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;
OC   Mammalia; Eutheria; Euarchontoglires; Primates; Catarrhini; Hominidae;
OC   Homo.
OX   NCBI_TaxID=9606;
RN   [1]
RP   NUCLEOTIDE SEQUENCE.
RC   TISSUE=Endometrium;
RG   The German cDNA Consortium;
RA   Poustka A., Albert R., Moosmayer P., Schupp I., Wellenreuther R.,
RA   Mewes H.W., Weil B., Amid C., Osanger A., Fobo G., Han M., Wiemann S.;
RL   Submitted (JAN-2005) to the EMBL/GenBank/DDBJ databases.
DR   EMBL; BX647768; CAI46039.1; -; mRNA.
KW   Hypothetical protein.
FT   NON_TER     408    408
SQ   SEQUENCE   408 AA;  45625 MW;  E3E36BFEA8B4284B CRC64;

  Query Match           64.1%;  Score 2148;  DB 2;  Length 408;
  Best Local Similarity   99.8%;  Pred. No. 2.4e-93;
  Matches  407;  Conservative    0;  Mismatches    1;  Indels    0;  Gaps    0;

Qy          1 MPQPSVSGMDPPFGDAFRSHTFSEQTLMSTDLLANSSDPDFMYELDREMNYQQNPRDNFL 60
              ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db          1 MPQPSVSGMDPPFGDAFRSHTFSEQTLMSTDLLANSSDPDFMYELDREMNYQQNPRDNFL 60

Qy         61 SLEDCKDIENLESFTDVLDNEGALTSNWEQWDTYCEDLTKYTKLTSCDIWGTKEVDYLGL 120
              ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db         61 SLEDCKDIENLESFTDVLDNEGALTSNWEQWDTYCEDLTKYTKLTSCDIWGTKEVDYLGL 120

Qy        121 DDFSSPYQDEEVISKTPTLAQLNSEDSQSVSDSLYYPDSLFSVKQNPLPSSFPGKKITSR 180
              ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db        121 DDFSSPYQDEEVISKTPTLAQLNSEDSQSVSDSLYYPDSLFSVKQNPLPSSFPGKKITSR 180

Qy        181 AAAPVCSSKTLQAEVPLSDCVQKASKPPSSTQIMVKTNMYHNEKVNFHVECKDYVKKAKV 240
              |||||||||||||||||||||||| |||||||||||||||||||||||||||||||||||
Db        181 AAAPVCSSKTLQAEVPLSDCVQKASKPTSSTQIMVKTNMYHNEKVNFHVECKDYVKKAKV 240

Qy        241 KINPVQQSRPLLSQIHTDAAKENTCYCGAVAKRQEKKGMEPLQGHATPALPFKETQELLL 300
              ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db        241 KINPVQQSRPLLSQIHTDAAKENTCYCGAVAKRQEKKGMEPLQGHATPALPFKETQELLL 300

Qy        301 SPLPQEGPGSLAAGESSSLSASTSVSDSSQKKEEHNYSLFVSDNLGEQPTKCSPEEDEED 360
              ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db        301 SPLPQEGPGSLAAGESSSLSASTSVSDSSQKKEEHNYSLFVSDNLGEQPTKCSPEEDEED 360

Qy        361 EEDVDDEDHDEGFGSEHELSENEEEEEEEEDYEDDKDDDISDTFSEPG 408
              |||||||||||||||||||||||||||||||||||||||||||||||||
Db        361 EEDVDDEDHDEGFGSEHELSENEEEEEEEEDYEDDKDDDISDTFSEPG 408
```

```
RESULT 6
Q86YR3_HUMAN
ID   Q86YR3_HUMAN PRELIMINARY;      PRT;   417 AA.
AC   Q86YR3;
DT   01-JUN-2003 (TrEMBLrel. 24, Created)
DT   01-JUN-2003 (TrEMBLrel. 24, Last sequence update)
DT   01-JUN-2003 (TrEMBLrel. 24, Last annotation update)
DE   Hypothetical protein.
OS   Homo sapiens (Human).
OC   Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;
OC   Mammalia; Eutheria; Euarchontoglires; Primates; Catarrhini; Hominidae;
OC   Homo.
OX   NCBI_TaxID=9606;
RN   [1]
RP   NUCLEOTIDE SEQUENCE.
RA   Li F., Yao K.T.;
RL   Submitted (NOV-2002) to the EMBL/GenBank/DDBJ databases.
DR   EMBL; AY174896; AAO18732.1; -; mRNA.
KW   Hypothetical protein.
SQ   SEQUENCE   417 AA;  46579 MW;  9DBD37B07C14556B CRC64;

  Query Match           64.1%;  Score 2148;  DB 2;  Length 417;
  Best Local Similarity   99.8%;  Pred. No. 2.4e-93;
  Matches  407;  Conservative    0;  Mismatches    1;  Indels    0;  Gaps    0;

Qy          1 MPQPSVSGMDPPFGDAFRSHTFSEQTLMSTDLLANSSDPDFMYELDREMNYQQNPRDNFL 60
              ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db          1 MPQPSVSGMDPPFGDAFRSHTFSEQTLMSTDLLANSSDPDFMYELDREMNYQQNPRDNFL 60

Qy         61 SLEDCKDIENLESFTDVLDNEGALTSNWEQWDTYCEDLTKYTKLTSCDIWGTKEVDYLGL 120
              ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db         61 SLEDCKDIENLESFTDVLDNEGALTSNWEQWDTYCEDLTKYTKLTSCDIWGTKEVDYLGL 120

Qy        121 DDFSSPYQDEEVISKTPTLAQLNSEDSQSVSDSLYYPDSLFSVKQNPLPSSFPGKKITSR 180
              ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db        121 DDFSSPYQDEEVISKTPTLAQLNSEDSQSVSDSLYYPDSLFSVKQNPLPSSFPGKKITSR 180
```

```
Qy        181 AAAPVCSSKTLQAEVPLSDCVQKASKPPSSTQIMVKTNMYHNEKVNFHVECKDYVKKAKV 240
              ||||||||||||||||||||||||||||| |||||||||||||||||||||||||||||||
Db        181 AAAPVCSSKTLQAEVPLSDCVQKASKPTSSTQIMVKTNMYHNEKVNFHVECKDYVKKAKV 240

Qy        241 KINPVQQSRPLLSQIHTDAAKENTCYCGAVAKRQEKKGMEPLQGHATPALPFKETQELLL 300
              ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db        241 KINPVQQSRPLLSQIHTDAAKENTCYCGAVAKRQEKKGMEPLQGHATPALPFKETQELLL 300

Qy        301 SPLPQEGPGSLAAGESSSLSASTSVSDSSQKKEEHNYSLFVSDNLGEQPTKCSPEEDEED 360
              ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db        301 SPLPQEGPGSLAAGESSSLSASTSVSDSSQKKEEHNYSLFVSDNLGEQPTKCSPEEDEED 360

Qy        361 EEDVDDEDHDEGFGSEHELSENEEEEEEEEDYEDDKDDDISDTFSEPG 408
              ||||||||||||||||||||||||||||||||||||||||||||||||
Db        361 EEDVDDEDHDEGFGSEHELSENEEEEEEEEDYEDDKDDDISDTFSEPG 408
```

RESULT 7
Q8C8N7_MOUSE
ID Q8C8N7_MOUSE PRELIMINARY; PRT; 351 AA.
AC Q8C8N7;
DT 01-MAR-2003 (TrEMBLrel. 23, Created)
DT 01-MAR-2003 (TrEMBLrel. 23, Last sequence update)
DT 01-JUN-2003 (TrEMBLrel. 24, Last annotation update)
DE Mus musculus adult retina cDNA, RIKEN full-length enriched library,
DE clone:A930040G19 product:hypothetical protein, full insert sequence.
DE (Fragment).
GN Name=A930001N09Rik;
OS Mus musculus (Mouse).
OC Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;
OC Mammalia; Eutheria; Euarchontoglires; Glires; Rodentia; Sciurognathi;
OC Muroidea; Muridae; Murinae; Mus.
OX NCBI_TaxID=10090;
RN [1]
RP NUCLEOTIDE SEQUENCE.
RC STRAIN=C57BL/6J; TISSUE=Retina;
RX MEDLINE=99279253; PubMed=10349636; DOI=10.1016/S0076-6879(99)03004-9;
RA Carninci P., Hayashizaki Y.;
RT "High-efficiency full-length cDNA cloning.";
RL Meth. Enzymol. 303:19-44(1999).
RN [2]
RP NUCLEOTIDE SEQUENCE.
RC STRAIN=C57BL/6J; TISSUE=Retina;
RX MEDLINE=21085660; PubMed=11217851; DOI=10.1038/35055500;
RA Kawai J., Shinagawa A., Shibata K., Yoshino M., Itoh M., Ishii Y.,
RA Arakawa T., Hara A., Fukunishi Y., Konno H., Adachi J., Fukuda S.,
RA Aizawa K., Izawa M., Nishi K., Kiyosawa H., Kondo S., Yamanaka I.,
RA Saito T., Okazaki Y., Gojobori T., Bono H., Kasukawa T., Saito R.,
RA Kadota K., Matsuda H.A., Ashburner M., Batalov S., Casavant T.,
RA Fleischmann W., Gaasterland T., Gissi C., King B., Kochiwa H.,
RA Kuehl P., Lewis S., Matsuo Y., Nikaido I., Pesole G., Quackenbush J.,
RA Schriml L.M., Staubli F., Suzuki R., Tomita M., Wagner L., Washio T.,
RA Sakai K., Okido T., Furuno M., Aono H., Baldarelli R., Barsh G.,
RA Blake J., Boffelli D., Bojunga N., Carninci P., de Bonaldo M.F.,
RA Brownstein M.J., Bult C., Fletcher C., Fujita M., Gariboldi M.,
RA Gustincich S., Hill D., Hofmann M., Hume D.A., Kamiya M., Lee N.H.,
RA Lyons P., Marchionni L., Mashima J., Mazzarelli J., Mombaerts P.,
RA Nordone P., Ring B., Ringwald M., Rodriguez I., Sakamoto N.,
RA Sasaki H., Sato K., Schoenbach C., Seya T., Shibata Y., Storch K.-F.,
RA Suzuki H., Toyo-oka K., Wang K.H., Weitz C., Whittaker C., Wilming L.,
RA Wynshaw-Boris A., Yoshida K., Hasegawa Y., Kawaji H., Kohtsuki S.,
RA Hayashizaki Y.;
RT "Functional annotation of a full-length mouse cDNA collection.";
RL Nature 409:685-690(2001).
RN [3]
RP NUCLEOTIDE SEQUENCE.
RC STRAIN=C57BL/6J; TISSUE=Retina;
RA The FANTOM Consortium,
RA the RIKEN Genome Exploration Research Group Phase I & II Team;
RT "Analysis of the mouse transcriptome based on functional annotation of
RT 60,770 full-length cDNAs.";
RL Nature 420:563-573(2002).
RN [4]
RP NUCLEOTIDE SEQUENCE.
RC STRAIN=C57BL/6J; TISSUE=Retina;
RX MEDLINE=20499374; PubMed=11042159; DOI=10.1101/gr.145100;
RA Carninci P., Shibata Y., Hayatsu N., Sugahara Y., Shibata K., Itoh M.,
RA Konno H., Okazaki Y., Muramatsu M., Hayashizaki Y.;
RT "Normalization and subtraction of cap-trapper-selected cDNAs to
RT prepare full-length cDNA libraries for rapid discovery of new genes.";
RL Genome Res. 10:1617-1630(2000).
RN [5]
RP NUCLEOTIDE SEQUENCE.
RC STRAIN=C57BL/6J; TISSUE=Retina;
RX MEDLINE=20530913; PubMed=11076861; DOI=10.1101/gr.152600;
RA Shibata K., Itoh M., Aizawa K., Nagaoka S., Sasaki N., Carninci P.,
RA Konno H., Akiyama J., Nishi K., Kitsunai T., Tashiro H., Itoh M.,
RA Sumi N., Ishii Y., Nakamura S., Hazama M., Nishine T., Harada A.,
RA Yamamoto R., Matsumoto H., Sakaguchi S., Ikegami T., Kashiwagi K.,
RA Fujiwake S., Inoue K., Togawa Y., Izawa M., Ohara E., Watahiki M.,
RA Yoneda Y., Ishikawa T., Ozawa K., Tanaka T., Matsuura S., Kawai J.,
RA Okazaki Y., Muramatsu M., Inoue Y., Kira A., Hayashizaki Y.;
RT "RIKEN integrated sequence analysis (RISA) system-384-format
RT sequencing pipeline with 384 multicapillary sequencer.";
RL Genome Res. 10:1757-1771(2000).
RN [6]
RP NUCLEOTIDE SEQUENCE.
RC STRAIN=C57BL/6J; TISSUE=Retina;
RA Adachi J., Aizawa K., Akimura T., Arakawa T., Bono H., Carninci P.,
RA Fukuda S., Furuno M., Hanagaki T., Hara A., Hashizume W.,
RA Hayashida K., Hayatsu N., Hiramoto K., Hiraoka T., Hirozane T.,
RA Hori F., Imotani K., Ishii Y., Itoh M., Kagawa I., Kasukawa T.,
RA Katoh H., Kawai J., Kojima Y., Kondo S., Konno H., Kouda M., Koya S.,
RA Kurihara C., Matsuyama T., Miyazaki A., Murata M., Nakamura M.,
RA Nishi K., Nomura K., Numazaki R., Ohno M., Ohsato N., Okazaki Y.,
RA Saito R., Saitoh H., Sakai C., Sakai K., Sakazume N., Sano H.,
RA Sasaki D., Shibata K., Shinagawa A., Shiraki T., Sogabe Y., Tagami M.,
RA Tagawa A., Takahashi F., Takaku-Akahira S., Takeda Y., Tanaka T.,
RA Tomaru A., Toya T., Yasunishi A., Muramatsu M., Hayashizaki Y.;
RL Submitted (JUL-2001) to the EMBL/GenBank/DDBJ databases.
DR EMBL; AK044760; BAC32070.1; -; mRNA.
DR Ensembl; ENSMUSG00000048249; Mus musculus.
DR MGI; MGI:1924378; A930001N09Rik.
KW Hypothetical protein.
FT NON_TER 351 351
SQ SEQUENCE 351 AA; 39254 MW; ABA4B36D996AFAAD CRC64;

  Query Match 50.9%; Score 1704.5; DB 2; Length 351;
  Best Local Similarity 92.6%; Pred. No. 1.4e-72;
  Matches 325; Conservative 9; Mismatches 16; Indels 1; Gaps 1;

```
Qy          1 MPQPSVSGMDPPFGDAFRSHTFSEQTLMSTDLLANSSDPDFMYELDREMNYQQNPRDNFL 60
              ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db          1 MPQPSVSGMDPPFGDAFRSHTFSEQTLMSTDLLANSSDPDFMYELDREMNYQQNPRDNFL 60

Qy         61 SLEDCKDIENLESFTDVLDNEGALTSNWEQWDTYCEDLTKYTKLTSCDIWGTKEVDYLGL 120
              ||||||||||||:|||||||| ||||||||||||||||||||||||||||||||||||||||
Db         61 SLEDCKDIENLETFTDVLDNEDALTSNWEQWDTYCEDLTKYTKLTSCDIWGTKEVDYLGL 120

Qy        121 DDFSSPYQDEEVISKTPTLAQLNSEDSQSVSDSLYYPDSLFSVKQNPL-PSSFPGKKITS 179
              |||||||||||||||||||||||||||||||||||||||||||||||| ||||| ||||:
Db        121 DDFSSPYQDEEVISKTPTLAQLNSEDSQSVSDSLYYPDSLFSVKQNPLPPSSFPSKKITN 180

Qy        180 RAAAPVCSSKTLQAEVPLSDCVQKASKPPSSTQIMVKTNMYHNEKVNFHVECKDYVKKAK 239
              |||||||||||||||||| |||||||||| ||||||||||||||||||||||||||||||||
Db        181 RAAAPVCSSKTLQAEVPSSDCVQKASKPTSSTQIMVKTNMYHNEKVNFHVECKDYVKKAK 240

Qy        240 VKINPVQQSRPLLSQIHTDAAKENTCYCGAVAKRQEKKGMEPLQGHATPALPFKETQELL 299
              |||||||| |||||:| |||||||||||||||||::|:|| ||  ||||||||||||||
Db        241 VKINPVQQGRPLLSQVHIDAAKENTCYCGAVAKRQERRGVEPHQGRGTPALPFKETQELL 300

Qy        300 LSPLPQEGPGSLAAGESSSLSASTSVSDSSQKKEEHNYSLFVSDNLGEQPT 350
              |||| |: || :|  || ||||||||||||||||||||||: ||||
Db        301 LSPLTQDSPGLVATAESGSLSASTSVSDSSQKKEEHNYSLFVSDNMREQPT 351
```

```
RESULT 8
Q9CTQ7_MOUSE
ID   Q9CTQ7_MOUSE PRELIMINARY;      PRT;   330 AA.
AC   Q9CTQ7;
DT   01-JUN-2001 (TrEMBLrel. 17, Created)
DT   01-JUN-2001 (TrEMBLrel. 17, Last sequence update)
DT   01-JUN-2001 (TrEMBLrel. 17, Last annotation update)
DE   Mus musculus adult retina cDNA, RIKEN full-length enriched library,
DE   clone:A930001N09 product:hypothetical protein, full insert sequence.
DE   (Fragment).
GN   Name=A930001N09Rik;
OS   Mus musculus (Mouse).
OC   Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;
OC   Mammalia; Eutheria; Euarchontoglires; Glires; Rodentia; Sciurognathi;
OC   Muroidea; Muridae; Murinae; Mus.
OX   NCBI_TaxID=10090;
RN   [1]
RP   NUCLEOTIDE SEQUENCE.
RC   STRAIN=C57BL/6J; TISSUE=Retina;
RX   MEDLINE=99279253; PubMed=10349636; DOI=10.1016/S0076-6879(99)03004-9;
RA   Carninci P., Hayashizaki Y.;
RT   "High-efficiency full-length cDNA cloning.";
RL   Meth. Enzymol. 303:19-44(1999).
RN   [2]
RP   NUCLEOTIDE SEQUENCE.
RC   STRAIN=C57BL/6J; TISSUE=Retina;
RX   MEDLINE=21085660; PubMed=11217851; DOI=10.1038/35055500;
RA   Kawai J., Shinagawa A., Shibata K., Yoshino M., Itoh M., Ishii Y.,
RA   Arakawa T., Hara A., Fukunishi Y., Konno H., Adachi J., Fukuda S.,
RA   Aizawa K., Izawa M., Nishi K., Kiyosawa H., Kondo S., Yamanaka I.,
RA   Saito T., Okazaki Y., Gojobori T., Bono H., Kasukawa T., Saito R.,
RA   Kadota K., Matsuda H.A., Ashburner M., Batalov S., Casavant T.,
RA   Fleischmann W., Gaasterland T., Gissi C., King B., Kochiwa H.,
RA   Kuehl P., Lewis S., Matsuo Y., Nikaido I., Pesole G., Quackenbush J.,
RA   Schriml L.M., Staubli F., Suzuki R., Tomita M., Wagner L., Washio T.,
RA   Sakai K., Okido T., Furuno M., Aono H., Baldarelli R., Barsh G.,
RA   Blake J., Boffelli D., Bojunga N., Carninci P., de Bonaldo M.F.,
RA   Brownstein M.J., Bult C., Fletcher C., Fujita M., Gariboldi M.,
RA   Gustincich S., Hill D., Hofmann M., Hume D.A., Kamiya M., Lee N.H.,
RA   Lyons P., Marchionni L., Mashima J., Mazzarelli J., Mombaerts P.,
RA   Nordone P., Ring B., Ringwald M., Rodriguez I., Sakamoto N.,
RA   Sasaki H., Sato K., Schoenbach C., Seya T., Shibata Y., Storch K.-F.,
RA   Suzuki H., Toyo-oka K., Wang K.H., Weitz C., Whittaker C., Wilming L.,
RA   Wynshaw-Boris A., Yoshida K., Hasegawa Y., Kawaji H., Kohtsuki S.,
RA   Hayashizaki Y.;
RT   "Functional annotation of a full-length mouse cDNA collection.";
RL   Nature 409:685-690(2001).
RN   [3]
RP   NUCLEOTIDE SEQUENCE.
RC   STRAIN=C57BL/6J; TISSUE=Retina;
RA   The FANTOM Consortium,
RA   the RIKEN Genome Exploration Research Group Phase I & II Team;
RT   "Analysis of the mouse transcriptome based on functional annotation of
RT   60,770 full-length cDNAs.";
RL   Nature 420:563-573(2002).
RN   [4]
RP   NUCLEOTIDE SEQUENCE.
RC   STRAIN=C57BL/6J; TISSUE=Retina;
RX   MEDLINE=20499374; PubMed=11042159; DOI=10.1101/gr.145100;
RA   Carninci P., Shibata Y., Hayatsu N., Sugahara Y., Shibata K., Itoh M.,
RA   Konno H., Okazaki Y., Muramatsu M., Hayashizaki Y.;
RT   "Normalization and subtraction of cap-trapper-selected cDNAs to
RT   prepare full-length cDNA libraries for rapid discovery of new genes.";
RL   Genome Res. 10:1617-1630(2000).
RN   [5]
RP   NUCLEOTIDE SEQUENCE.
RC   STRAIN=C57BL/6J; TISSUE=Retina;
RX   MEDLINE=20530913; PubMed=11076861; DOI=10.1101/gr.152600;
RA   Shibata K., Itoh M., Aizawa K., Nagaoka S., Sasaki N., Carninci P.,
RA   Konno H., Akiyama J., Nishi K., Kitsunai T., Tashiro H., Itoh M.,
RA   Sumi N., Ishii Y., Nakamura S., Hazama M., Nishine T., Harada A.,
RA   Yamamoto R., Matsumoto H., Sakaguchi S., Ikegami T., Kashiwagi K.,
RA   Fujiwake S., Inoue K., Togawa Y., Izawa M., Ohara E., Watahiki M.,
RA   Yoneda Y., Ishikawa T., Ozawa K., Tanaka T., Matsuura S., Kawai J.,
RA   Okazaki Y., Muramatsu M., Inoue Y., Kira A., Hayashizaki Y.;
RT   "RIKEN integrated sequence analysis (RISA) system-384-format
RT   sequencing pipeline with 384 multicapillary sequencer.";
RL   Genome Res. 10:1757-1771(2000).
RN   [6]
RP   NUCLEOTIDE SEQUENCE.
RC   STRAIN=C57BL/6J; TISSUE=Retina;
RA   Adachi J., Aizawa K., Akahira S., Akimura T., Arai A., Aono H.,
RA   Arakawa T., Bono H., Carninci P., Fukuda S., Fukunishi Y., Furuno M.,
RA   Hanagaki T., Hara A., Hayatsu N., Hiramoto K., Hiraoka T., Hori F.,
RA   Imotani K., Ishii Y., Itoh M., Izawa M., Kasukawa T., Kato H.,
RA   Kawai J., Kojima Y., Konno H., Kouda M., Koya S., Kurihara C.,
RA   Matsuyama T., Miyazaki A., Nishi K., Nomura K., Numazaki R., Ohno M.,
RA   Okazaki Y., Okido T., Owa C., Saito H., Saito R., Sakai C., Sakai K.,
RA   Sano H., Sasaki D., Shibata K., Shibata Y., Shinagawa A., Shiraki T.,
RA   Sogabe Y., Suzuki H., Tagami M., Tagawa A., Takahashi F., Tanaka T.,
RA   Tejima Y., Toya T., Yamamura T., Yasunishi A., Yoshida K., Yoshino M.,
RA   Muramatsu M., Hayashizaki Y.;
RL   Submitted (AUG-2000) to the EMBL/GenBank/DDBJ databases.
DR   EMBL; AK020796; BAB32214.1; -; mRNA.
DR   Ensembl; ENSMUSG00000048249; Mus musculus.
DR   MGI; MGI:1924378; A930001N09Rik.
KW   Hypothetical protein.
FT   NON_TER     330    330
SQ   SEQUENCE   330 AA;  36692 MW;  1397555C4934A64B CRC64;

  Query Match           47.9%;  Score 1602.5;  DB 2;  Length 330;
  Best Local Similarity  92.7%;  Pred. No. 8e-68;
  Matches  306;  Conservative    8;  Mismatches   15;  Indels    1;  Gaps    1;

Qy          1 MPQPSVSGMDPPFGDAFRSHTFSEQTLMSTDLLANSSDPDFMYELDREMNYQQNPRDNFL 60
              ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db          1 MPQPSVSGMDPPFGDAFRSHTFSEQTLMSTDLLANSSDPDFMYELDREMNYQQNPRDNFL 60

Qy         61 SLEDCKDIENLESFTDVLDNEGALTSNWEQWDTYCEDLTKYTKLTSCDIWGTKEVDYLGL 120
              ||||||||||||:|||||||| ||||||||||||||||||||||||||||||||||||||
Db         61 SLEDCKDIENLETFTDVLDNEDALTSNWEQWDTYCEDLTKYTKLTSCDIWGTKEVDYLGL 120

Qy        121 DDFSSPYQDEEVISKTPTLAQLNSEDSQSVSDSLYYPDSLFSVKQNPL-PSSFPGKKITS 179
              |||||||||||||||||||||||||||||||||||||||||||||||| ||||| ||||:
Db        121 DDFSSPYQDEEVISKTPTLAQLNSEDSQSVSDSLYYPDSLFSVKQNPLPPSSFPSKKITN 180

Qy        180 RAAAPVCSSKTLQAEVPLSDCVQKASKPPSSTQIMVKTNMYHNEKVNFHVECKDYVKKAK 239
              |||||||||||||||| |||||||||| ||||||||||||||||||||||||||||||||
Db        181 RAAAPVCSSKTLQAEVPSSDCVQKASKPTSSTQIMVKTNMYHNEKVNFHVECKDYVKKAK 240

Qy        240 VKINPVQQSRPLLSQIHTDAAKENTCYCGAVAKRQEKKGMEPLQGHATPALPFKETQELL 299
              |||||||| |||||:| |||||||||||||||||::|:|| ||  ||||||||||||
Db        241 VKINPVQQGRPLLSQVHIDAAKENTCYCGAVAKRQERRGVEPHQGRGTPALPFKETQELL 300

Qy        300 LSPLPQEGPGSLAAGESSSLSASTSVSDSS 329
              |||| |: || :|  || ||||||||||||
Db        301 LSPLTQDSPGLVATAESGSLSASTSVSDSS 330


RESULT 9
Q4RRX3_TETNG
ID   Q4RRX3_TETNG PRELIMINARY;      PRT;   600 AA.
AC   Q4RRX3;
DT   13-SEP-2005 (TrEMBLrel. 31, Created)
DT   13-SEP-2005 (TrEMBLrel. 31, Last sequence update)
DT   13-SEP-2005 (TrEMBLrel. 31, Last annotation update)
DE   Chromosome 7 SCAF15001, whole genome shotgun sequence.
DE   (Fragment).
GN   ORFNames=GSTENG00029962001;
OS   Tetraodon nigroviridis (Green puffer).
OC   Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;
OC   Actinopterygii; Neopterygii; Teleostei; Euteleostei; Neoteleostei;
OC   Acanthomorpha; Acanthopterygii; Percomorpha; Tetraodontiformes;
OC   Tetradontoidea; Tetraodontidae; Tetraodon.
OX   NCBI_TaxID=99883;
RN   [1]
```

```
RP   NUCLEOTIDE SEQUENCE.
RA   Jaillon O., Aury J.M., Brunet F., Petit J.L., Stange-Thomann N.,
RA   Mauceli E., Bouneau L., Fischer C., Ozouf-Costaz C., Bernot A.,
RA   Nicaud S., Jaffe D., Fisher S., Lutfalla G., Dossat C., Segurens B.,
RA   Dasilva C., Salanoubat M., Levy M., Boudet N., Castellano S.,
RA   Anthouard V., Jubin C., Castelli V., Katinka M., Vacherie B.,
RA   Biemont C., Skalli Z., Cattolico L., Poulain J., De Berardinis V.,
RA   Cruaud C., Duprat S., Brottier P., Coutanceau J.P., Gouzy J.,
RA   Parra G., Lardier G., Chapple C., McKernan K.J., McEwan P., Bosak S.,
RA   Kellis M., Volff JN., Guigo R., Zody M.C., Mesirov J.,
RA   Lindblad-Toh K., Birren B., Nusbaum C., Kahn D., Robinson-Rechavi M.,
RA   Laudet V., Schachter V., Quetier F., Saurin W., Scarpelli C.,
RA   Wincker P., Lander E.S., Weissenbach J., Roest Crollius H.;
RT   "Genome duplication in the teleost fish Tetraodon nigroviridis reveals
RT   the early vertebrate proto-karyotype.";
RL   Nature 431:946-957(2004).
RN   [2]
RP   NUCLEOTIDE SEQUENCE.
RG   Genoscope; Whitehead Institute Centre for Genome Research;
RL   Submitted (FEB-2004) to the EMBL/GenBank/DDBJ databases.
CC   -!- CAUTION: The sequence shown here is derived from an
CC       EMBL/GenBank/DDBJ whole genome shotgun (WGS) entry which is
CC       preliminary data.
DR   EMBL; CAAE01015001; CAG08859.1; -; Genomic_DNA.
FT   NON_TER       1      1
FT   NON_TER     600    600
SQ   SEQUENCE   600 AA;  66291 MW;  B2F28B2F3B27B1C4 CRC64;

  Query Match             41.1%;  Score 1377.5;  DB 2;  Length 600;
  Best Local Similarity   43.8%;  Pred. No. 5.9e-57;
  Matches  316;  Conservative   74;  Mismatches  126;  Indels  205;  Gaps   15;

Qy          3 QPSVSGMDPPFGDAFRSHTFSEQTLMSTDLLANSSDPDFMYELDREMNYQQNP------- 55
              |||||||:|||||||::::| :| | ||:||| ||||||||||||: |:||:|
Db          1 QPSVSGMEPPFGDAFQNYSFVDQALTSTELLATSSDPDFMYELDRDTNHQQSPCGDSMVG 60

Qy         56 -----------RDNFLSLEDCKDIENLESFTDVLDNEGALTSNWEQWDTYCEDLTKYTKL 104
                        |  : | :|: : :  :|              ||||:| ||||:||:|
Db         61 VGDGGKEVEGCVDQLMGLGECETVHSSSAF-------------EQWDSYWEDLTRYTRL 106

Qy        105 TSCDIWGTKEVDYLGLDDFSSPYQDEEVISKTPTLAQLNSEDSQSVSDSLYYPDSLFSVK 164
              |||||||||||:|||||||||||||||| :||||||||||||  | ::|| |  |
Db        107 ASCDIWGTKEVDFLGLDDFSSPYQDEEVIGRTPTLAQLNSEDSLPVCEALYPPADL---- 162

Qy        165 QNPLPSSFPGKKITSRAAAPVCSSK----------------TLQAEVPLSDCV---QKA 204
                    | || :  |:     | ||                    :  :  || |     |||
Db        163 ------SLPGPQPQSQPPQLPCHSKRPVVPGSVRSSSSSTSSSRPSRSPLPDFAEGSQKA 216

Qy        205 SKP-PSSTQIMVKTNMYHNEKVNFHVECKDYVK-KAKVKINPVQQSRPLLSQIHTDAAKE 262
              ::| ||||: | |        |     :|| : : |::    : : |  :  ::|
Db        217 TRPVPSSTETMAKNQ-------NLFSLSQDYGQGQPKLQARATKMAAP--APHNSDFVPP 267

Qy        263 NTCYCGAVAKRQEKKGMEPLQGHATPALP-FKETQELLLSPLPQEGPGSLAAGESSSLSA 321
              :    :  :| |  |   :    : |:|   |   :|
Db        268 ASSNLMSFERRVEATGRTDVPAGWSSAVPHLVEANRVL---------------------- 305

Qy        322 STSVSDSSQKKEEHNYSLFVSDNLGEQPTKCSPEEDEEDEEDVDDEDHDEGFGSEHELSE 381
                    | | |   |:::|     |    | |
Db        306 -----------EVNASGLASESVGAAEAICGGSCDME--------------------- 331

Qy        382 NEEEEEEEEDYEDDKDDDISDTFSEPGYENDSVEDLKEVTS-ISSRKRGKRRYFWEYSEQ 440
                                          :||:| :|: :|||||||||||||||||
Db        332 -----------------------------LMEDIKGLTAGVSSRKRGKRRYFWEYSEQ 360

Qy        441 LTPSQQERMLRPSEWNRDTLPSNMYQKNGLHHGKYAVKKSRRTDVEDLTPNPKKLLQIGN 500
              | ||:|||||:||||:| |||||:|||||  |||| :|||||||||||||:|||||||
Db        361 LPPSKQERMLKPSEWDRHTLPSNLYQKNGPLHGKYMLKKSRRTDVEDLTPNPRKLLQIGT 420

Qy        501 ELRKLNKVISDLTPVSELPLTARPRSRKEKNKLAFRACRLKKKAQYEANKVKLWGLNTEY 560
              |||||||||||||||||||||||||||||||| ||||||||||||||||||||:|||
Db        421 ELRKLNKVISDLTPVSELPLTARPRSRKEKNKLASRACRLKKKAQYEANKVKLWGLSTEY 480

Qy        561 DNLLFVINSIKQEIVNRVQNPRDERGPNMGQKLEILIKDTLG------------------ 602
```

```
              | ||||||:||:||| ||::    |  ||   || ||: |||
Db        481 DRLLFVINTIKEEIVARVED-SSPRPTNMTDTLEHLIQQTLGENPFRRDSQQNQTIFSQK 539

Qy        603 -----------------------LPVAGQTSEFVNQVLEKTAEGNPTGGLVGLRIPTSK 638
                                    |||||||:|||::|: |  |:||||||||:||||
Db        540 RKVNDVNVDLKYILKHFSLSPQVAPPVAGQTSDFVNKILDNTGRGDPTGGLVGLRVPTSK 599

Qy        639 V 639
              :
Db        600 I 600
```

RESULT 10
Q5ZMD5_CHICK
ID   Q5ZMD5_CHICK PRELIMINARY;   PRT;   131 AA.
AC   Q5ZMD5;
DT   25-OCT-2004 (TrEMBLrel. 28, Created)
DT   25-OCT-2004 (TrEMBLrel. 28, Last sequence update)
DT   25-OCT-2004 (TrEMBLrel. 28, Last annotation update)
DE   Hypothetical protein.
GN   ORFNames=RCJMB04_2h14;
OS   Gallus gallus (Chicken).
OC   Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;
OC   Archosauria; Aves; Neognathae; Galliformes; Phasianidae; Phasianinae;
OC   Gallus.
OX   NCBI_TaxID=9031;
RN   [1]
RP   NUCLEOTIDE SEQUENCE.
RC   STRAIN=CB; TISSUE=Bursa;
RA   Caldwell R.B., Kierzek A.M., Arakawa H., Bezzubov Y., Zaim J.,
RA   Fiedler P., Kutter S., Blagodatski A., Kostovska D., Koter M.,
RA   Plachy J., Carninci P., Hayashizaki Y., Buerstedde J.M.;
RT   "Full-length cDNAs from chicken bursal lymphocytes to facilitate
RT   genefunction analysis.";
RL   Genome Biol. 6:R6-R6(2005).
DR   EMBL; AJ719449; CAG31108.1; -; mRNA.
KW   Hypothetical protein.
SQ   SEQUENCE   131 AA; 15183 MW; 4D73D84C28DE3670 CRC64;

  Query Match   19.6%; Score 656; DB 2; Length 131;
  Best Local Similarity   90.1%; Pred. No. 8.5e-24;
  Matches 118; Conservative   6; Mismatches   7; Indels   0; Gaps   0;

Qy   1 MPQPSVSGMDPPFGDAFRSHTFSEQTLMSTDLLANSSDPDFMYELDREMNYQQNPRDNFL 60
              |||||||||||||||||||| |||||||||||||:|||||||||||||||:|||: ||| |
Db   1 MPQPSVSGMDPPFGDAFRSHVFSEQTLMSTDLLASSSDPDFMYELDREMDYQQSSRDNLL 60

Qy   61 SLEDCKDIENLESFTDVLDNEGALTSNWEQWDTYCEDLTKYTKLTSCDIWGTKEVDYLGL 120
              |:|||||:|||||||:|| | | || ||||||||||||||||||||||||||||||||||
Db   61 SMEDCKDLENLESFTDILDKEAAFTSKWEQWDTYCEDLTKYTKLTSCDIWGTKEVDYLGL 120

Qy   121 DDFSSPYQDEE 131
              ||||||||||:
Db   121 DDFSSPYQDEE 131

RESULT 11
Q9VC61_DROME
ID   Q9VC61_DROME PRELIMINARY;   PRT;   755 AA.
AC   Q9VC61; Q8T9A9;
DT   01-MAY-2000 (TrEMBLrel. 13, Created)
DT   01-MAR-2003 (TrEMBLrel. 23, Last sequence update)
DT   10-MAY-2005 (TrEMBLrel. 30, Last annotation update)
DE   CG13624-PA, isoform A (Cg13624-pb, isoform b) (SD09792p).
GN   ORFNames=CG13624;
OS   Drosophila melanogaster (Fruit fly).
OC   Eukaryota; Metazoa; Arthropoda; Hexapoda; Insecta; Pterygota;
OC   Neoptera; Endopterygota; Diptera; Brachycera; Muscomorpha;
OC   Ephydroidea; Drosophilidae; Drosophila.
OX   NCBI_TaxID=7227;
RN   [1]
RP   NUCLEOTIDE SEQUENCE.
RX   MEDLINE=20196006; PubMed=10731132; DOI=10.1126/science.287.5461.2185;

```
RA   Adams M.D., Celniker S.E., Holt R.A., Evans C.A., Gocayne J.D.,
RA   Amanatides P.G., Scherer S.E., Li P.W., Hoskins R.A., Galle R.F.,
RA   George R.A., Lewis S.E., Richards S., Ashburner M., Henderson S.N.,
RA   Sutton G.G., Wortman J.R., Yandell M.D., Zhang Q., Chen L.X.,
RA   Brandon R.C., Rogers Y.-H.C., Blazej R.G., Champe M., Pfeiffer B.D.,
RA   Wan K.H., Doyle C., Baxter E.G., Helt G., Nelson C.R., Miklos G.L.G.,
RA   Abril J.F., Agbayani A., An H.-J., Andrews-Pfannkoch C., Baldwin D.,
RA   Ballew R.M., Basu A., Baxendale J., Bayraktaroglu L., Beasley E.M.,
RA   Beeson K.Y., Benos P.V., Berman B.P., Bhandari D., Bolshakov S.,
RA   Borkova D., Botchan M.R., Bouck J., Brokstein P., Brottier P.,
RA   Burtis K.C., Busam D.A., Butler H., Cadieu E., Center A., Chandra I.,
RA   Cherry J.M., Cawley S., Dahlke C., Davenport L.B., Davies P.,
RA   de Pablos B., Delcher A., Deng Z., Mays A.D., Dew I., Dietz S.M.,
RA   Dodson K., Doup L.E., Downes M., Dugan-Rocha S., Dunkov B.C., Dunn P.,
RA   Durbin K.J., Evangelista C.C., Ferraz C., Ferriera S., Fleischmann W.,
RA   Fosler C., Gabrielian A.E., Garg N.S., Gelbart W.M., Glasser K.,
RA   Glodek A., Gong F., Gorrell J.H., Gu Z., Guan P., Harris M.,
RA   Harris N.L., Harvey D.A., Heiman T.J., Hernandez J.R., Houck J.,
RA   Hostin D., Houston K.A., Howland T.J., Wei M.-H., Ibegwam C.,
RA   Jalali M., Kalush F., Karpen G.H., Ke Z., Kennison J.A., Ketchum K.A.,
RA   Kimmel B.E., Kodira C.D., Kraft C., Kravitz S., Kulp D., Lai Z.,
RA   Lasko P., Lei Y., Levitsky A.A., Li J.H., Li Z., Liang Y., Lin X.,
RA   Liu X., Mattei B., McIntosh T.C., McLeod M.P., McPherson D.,
RA   Merkulov G., Milshina N.V., Mobarry C., Morris J., Moshrefi A.,
RA   Mount S.M., Moy M., Murphy B., Murphy L., Muzny D.M., Nelson D.L.,
RA   Nelson D.R., Nelson K.A., Nixon K., Nusskern D.R., Pacleb J.M.,
RA   Palazzolo M., Pittman G.S., Pan S., Pollard J., Puri V., Reese M.G.,
RA   Reinert K., Remington K., Saunders R.D.C., Scheeler F., Shen H.,
RA   Shue B.C., Siden-Kiamos I., Simpson M., Skupski M.P., Smith T.,
RA   Spier E., Spradling A.C., Stapleton M., Strong R., Sun E.,
RA   Svirskas R., Tector C., Turner R., Venter E., Wang A.H., Wang X.,
RA   Wang Z.-Y., Wassarman D.A., Weinstock G.M., Weissenbach J.,
RA   Williams S.M., Woodage T., Worley K.C., Wu D., Yang S., Yao Q.A.,
RA   Ye J., Yeh R.-F., Zaveri J.S., Zhan M., Zhang G., Zhao Q., Zheng L.,
RA   Zheng X.H., Zhong F.N., Zhong W., Zhou X., Zhu S., Zhu X., Smith H.O.,
RA   Gibbs R.A., Myers E.W., Rubin G.M., Venter J.C.;
RT   "The genome sequence of Drosophila melanogaster.";
RL   Science 287:2185-2195(2000).
RN   [2]
RP   NUCLEOTIDE SEQUENCE.
RX   MEDLINE=22426065; PubMed=12537568;
RA   Celniker S.E., Wheeler D.A., Kronmiller B., Carlson J.W., Halpern A.,
RA   Patel S., Adams M., Champe M., Dugan S.P., Frise E., Hodgson A.,
RA   George R.A., Hoskins R.A., Laverty T., Muzny D.M., Nelson C.R.,
RA   Pacleb J.M., Park S., Pfeiffer B.D., Richards S., Sodergren E.J.,
RA   Svirskas R., Tabor P.E., Wan K., Stapleton M., Sutton G.G., Venter C.,
RA   Weinstock G., Scherer S.E., Myers E.W., Gibbs R.A., Rubin G.M.;
RT   "Finishing a whole-genome shotgun: release 3 of the Drosophila
RT   melanogaster euchromatic genome sequence.";
RL   Genome Biol. 3:RESEARCH0079-RESEARCH0079(2002).
RN   [3]
RP   NUCLEOTIDE SEQUENCE.
RX   MEDLINE=22426070; PubMed=12537573;
RA   Kaminker J.S., Bergman C.M., Kronmiller B., Carlson J.W., Svirskas R.,
RA   Patel S., Frise E., Wheeler D.A., Lewis S.E., Rubin G.M.,
RA   Ashburner M., Celniker S.E.;
RT   "The transposable elements of the Drosophila melanogaster euchromatin:
RT   a genomics perspective.";
RL   Genome Biol. 3:RESEARCH0084.1-RESEARCH0084.20(2002).
RN   [4]
RP   NUCLEOTIDE SEQUENCE.
RX   MEDLINE=22426069; PubMed=12537572;
RA   Misra S., Crosby M.A., Mungall C.J., Matthews B.B., Campbell K.S.,
RA   Hradecky P., Huang Y., Kaminker J.S., Millburn G.H., Prochnik S.E.,
RA   Smith C.D., Tupy J.L., Whitfield E.J., Bayraktaroglu L., Berman B.P.,
RA   Bettencourt B.R., Celniker S.E., de Grey A.D.N.J., Drysdale R.A.,
RA   Harris N.L., Richter J., Russo S., Schroeder A.J., Shu S.Q.,
RA   Stapleton M., Yamada C., Ashburner M., Gelbart W.M., Rubin G.M.,
RA   Lewis S.E.;
RT   "Annotation of the Drosophila melanogaster euchromatic genome: a
RT   systematic review.";
RL   Genome Biol. 3:RESEARCH0083.1-RESEARCH0083.22(2002).
RN   [5]
RP   NUCLEOTIDE SEQUENCE.
RG   Berkeley Drosophila Genome Project;
RA   Celniker S., Carlson J., Wan K., Pfeiffer B., Frise E., George R.,
RA   Hoskins R., Stapleton M., Pacleb J., Park S., Svirskas R., Smith E.,
RA   Yu C., Rubin G.;
RT   "Drosophila melanogaster release 4 sequence.";
RL   Submitted (MAR-2000) to the EMBL/GenBank/DDBJ databases.
RN   [6]
RP   NUCLEOTIDE SEQUENCE.
RG   FlyBase;
RL   Submitted (MAR-2005) to the EMBL/GenBank/DDBJ databases.
RN   [7]
RP   NUCLEOTIDE SEQUENCE.
RA   Stapleton M., Brokstein P., Hong L., Agbayani A., Carlson J.,
RA   Champe M., Chavez C., Dorsett V., Farfan D., Frise E., George R.,
RA   Gonzalez M., Guarin H., Li P., Liao G., Miranda A., Mungall C.J.,
RA   Nunoo J., Pacleb J., Paragas V., Park S., Phouanenavong S., Wan K.,
RA   Yu C., Lewis S.E., Rubin G.M., Celniker S.;
RL   Submitted (DEC-2001) to the EMBL/GenBank/DDBJ databases.
CC   -!- INTERACTION:
CC       Q9VL14:CG18619; NbExp=1; IntAct=EBI-150517, EBI-165049;
DR   EMBL; AE003748; AAF56314.2; -; Genomic_DNA.
DR   EMBL; AY069850; AAL39995.1; -; mRNA.
DR   IntAct; Q9VC61; -.
DR   Ensembl; CG13624; Drosophila melanogaster.
DR   FlyBase; FBgn0039209; CG13624.
DR   InterPro; IPR004827; TF_bZIP.
DR   PROSITE; PS00036; BZIP_BASIC; UNKNOWN_1.
SQ   SEQUENCE   755 AA;  82263 MW;  D851D42235FF10E5 CRC64;

  Query Match             12.5%;  Score 417;  DB 2;  Length 755;
  Best Local Similarity   24.1%;  Pred. No. 1.1e-11;
  Matches  160;  Conservative   81;  Mismatches  183;  Indels  240;  Gaps   23;

Qy          5 SVSGMDPPFGDAF---RSHTFSEQTLMSTDLLANSSDPDFMYELDREMNYQQNPRDNFL- 60
              | :|:    |  |    ||: |  :    ||  :: |:          |: |: |  | :
Db        205 SSNGIGSGSGGYFTPDMSHSLSLNVVSEQVLLQEATTPN-------ELLYEMTPNSNAMW 257

Qy         61 ---------------SLED------------CKDIENLE--SFTDVLDNEGALTSNWEQW 91
                           ||:|               |: :|    | ||: |
Db        258 SDISSAIIHTKHEPFSLDDDYIFPNDKAEIQAADLSDLNGGDFLDVIGN----------- 306

Qy         92 DTYCEDLTKYTKLTSCDIWGTKEVDYL------GLDDFSSP-----------YQDEEV-- 132
                 ||    | :       |: |::|       | |   :|           :|  :|
Db        307 ---IEDFLPQTAV-------TQSVNFLLSPQAQGQDALVAPPMELLQQQQQNHQQLQVGS 356

Qy        133 ---ISKTPTLAQLNSEDSQSVSDSLYYPDSLFSVKQNPLPSSF-PGKKITSRAAAPVCSS 188
                  :    ||:|    :| | |   |  :    :|  | :|| || : :      |
Db        357 LPQLQTLLTLSQQQQSNSSSTSPYEIYHSTPQKPQQQQLSASFSPGSQASQSPLTP---- 412

Qy        189 KTLQAEVPLSDCVQKASKPPSST----QIMVKTNMYHNEKVNFHVECKDYVKKAKVKINP 244
                              || :     |::   ||             :: :||     :|
Db        413 ---------------PPPPHANRPQYQMVKSRNM-----------QELIKKGFPMSSP 444

Qy        245 VQQSRPLLSQIHTDAAKENTCYCGAVAKRQEKKGMEPLQGHATPALPFKETQELLLSPLP 304
                ::|  :||                                             |  |||
Db        445 PERS--ILS---------------------------------------------QSAALSPGG 460

Qy        305 QEGPGSLAAGESSSLSASTSVSDSSQKKEEHNYSLFVSDNLGEQPTKCSPEEDEEDEEDV 364
                | || |:| |:: |  ||    |  ::    |   | ::   : :  :|
Db        461 SSGFGSSASGNSTTTSNQTS---GSAVRKSFGYQSAVENSQLSRLSSSAPTH-------- 509

Qy        365 DDEDHDEGFGSEH---------------ELSENEE----------------EEEEEEDY 392
                    | ||                 |:| |                  :::|:|
Db        510 --------LGLEHIWMRREPRQHLLSTGSLAEAESFSSLSTGSVLSPDGIDFSQDDEDDN 561

Qy        393 EDDKDDDISDTFSEPGYENDSVEDLKEVTSISSRKRGKRRYFWEYSEQLTPSQQERML-- 450
                :  |:  |  |: |   |  |       :  |  :|| |:||:|: |    : :|::
Db        562 SSENSDNYDDCSSDNGLSEDEDETRTSTPNHLSSSKGKERFFWQYNVQAKGPKGKRLVFQ 621

Qy        451 ----RPSEWNRDTLP----------SNMYQKNGLHHGKYAVKKSRRTDVEDLTPNPKKLL 496
                    |   | |            :|: :|          |:|: |  ||||| :||
Db        622 SKLEDPHVLNEVTDPVFSPTCSVRGIKVYKHSG---------KARKGDGNDLTPNARKLH 672
```

```
Qy        497 QIGNELRKLNKVISDLTPVSELPLTARPRSRKEKNKLAFRACRLKKKAQYEANKVKLWGL 556
              || || ||:: |:|:|||||||    ||:||||||| ||||||||||:||||:||:||
Db        673 NIGKELDKLSRTINDMTPVSELPFNVRPKSRKEKNKLASRACRLKKKAQHEANKIKLFGL 732

Qy        557 NTEY 560
                |:
Db        733 EIEH 736


RESULT 12
Q7QG21_ANOGA
ID   Q7QG21_ANOGA PRELIMINARY;      PRT;   293 AA.
AC   Q7QG21;
DT   01-MAR-2004 (TrEMBLrel. 26, Created)
DT   01-MAR-2004 (TrEMBLrel. 26, Last sequence update)
DT   01-MAR-2004 (TrEMBLrel. 26, Last annotation update)
DE   ENSANGP00000015170 (Fragment).
GN   ORFNames=ENSANGG00000012681;
OS   Anopheles gambiae str. PEST.
OC   Eukaryota; Metazoa; Arthropoda; Hexapoda; Insecta; Pterygota;
OC   Neoptera; Endopterygota; Diptera; Nematocera; Culicoidea; Culicidae;
OC   Anophelinae; Anopheles.
OX   NCBI_TaxID=180454;
RN   [1]
RP   NUCLEOTIDE SEQUENCE.
RC   STRAIN=PEST;
RG   The Anopheles gambiae Sequence Committee;
RT   "Anopheles gambiae re-annotation.";
RL   Submitted (APR-2002) to the EMBL/GenBank/DDBJ databases.
RN   [2]
RP   NUCLEOTIDE SEQUENCE.
RC   STRAIN=PEST;
RG   The Anopheles gambiae Sequence Committee;
RL   Submitted (APR-2004) to the EMBL/GenBank/DDBJ databases.
CC   -!- CAUTION: The sequence shown here is derived from an
CC       EMBL/GenBank/DDBJ whole genome shotgun (WGS) entry which is
CC       preliminary data.
DR   EMBL; AAAB01008844; EAA06118.2; -; Genomic_DNA.
DR   GO; GO:0005634; C:nucleus; IEA.
DR   GO; GO:0003677; F:DNA binding; IEA.
DR   InterPro; IPR004827; TF_bZIP.
DR   PROSITE; PS00036; BZIP_BASIC; UNKNOWN_1.
FT   NON_TER       1       1
FT   NON_TER     293     293
SQ   SEQUENCE   293 AA;  31535 MW;  0700521A3C593F97 CRC64;

  Query Match             11.5%;  Score 386;  DB 2;  Length 293;
  Best Local Similarity   36.3%;  Pred. No. 1e-10;
  Matches  102;  Conservative   38;  Mismatches   79;  Indels   62;  Gaps    9;

Qy        311 LAAGESS------------------SLSASTSVSDSSQ--KKEEHNYSLFVSDNLGEQP 349
              |||| :|                  | || | :|   |  :: |     |  : :| |
Db         26 LAAGTTSAAVSVGSNAGTSGTAMSRLSSSAPTHISGFEQIWQRREPRPHLLSTGSLAEAG 85

Qy        350 TKCSPEEDEEDEEDVDDEDHDEGFGSEHELSENEEEEEEEEDYEDDKDDDISDTFSEPGY 409
              :  |         :  |  ||||:          :: :  |||   :| |  ||
Db         86 STSSLSTGSILSPEAPDFSHDEGY---------SDDSDHYEDYSSGEDSDSEDT------ 130

Qy        410 ENDSVEDLKEVTSISSRKRGKRRYFWEYSEQLTPSQQERML------RPSEWNRDTLPSN 463
                      :|:| :  |   :||||:|: |    : :|::       |   | | |
Db        131 --------GKVSSSAGSK--AKRYFWQYNVQAKGPKGQRLVIKTQVEDPHVLNAVTDP-- 178

Qy        464 MYQKN----GLHHGKYAVKKSRRTDVEDLTPNPKKLLQIGNELRKLNKVISDLTPVSELP 519
              ::  |    |: |       |:|: |  ||||||:||  || || || :||:|: ||||||
Db        179 VFSPNCSVRGIKHS----GKARKGDGNDLTPNPRKLHNIGKELDKLGRVINDMKPVSELP 234

Qy        520 LTARPRSRKEKNKLAFRACRLKKKAQYEANKVKLWGLNTEY 560
                |||::|||||||| ||||||||||:||||:||:||  ||:
Db        235 FNARPKTRKEKNKLASRACRLKKKAQHEANKIKLYGLETEH 275


RESULT 13
```

```
Q7RAS7_PLAYO
ID   Q7RAS7_PLAYO PRELIMINARY;      PRT;  2649 AA.
AC   Q7RAS7;
DT   01-MAR-2004 (TrEMBLrel. 26, Created)
DT   01-MAR-2004 (TrEMBLrel. 26, Last sequence update)
DT   01-MAR-2004 (TrEMBLrel. 26, Last annotation update)
DE   Hypothetical protein.
GN   Name=PY06422;
OS   Plasmodium yoelii yoelii.
OC   Eukaryota; Alveolata; Apicomplexa; Haemosporida; Plasmodium.
OX   NCBI_TaxID=73239;
RN   [1]
RP   NUCLEOTIDE SEQUENCE.
RC   STRAIN=17XNL;
RX   MEDLINE=22255706; PubMed=12368865; DOI=10.1038/nature01099;
RA   Carlton J.M., Angiuoli S.V., Suh B.B., Kooij T.W., Pertea M.,
RA   Silva J.C., Ermolaeva M.D., Allen J.E., Selengut J.D., Koo H.L.,
RA   Peterson J.D., Pop M., Kosack D.S., Shumway M.F., Bidwell S.L.,
RA   Shallom S.J., van Aken S.E., Riedmuller S.B., Feldblyum T.V.,
RA   Cho J.K., Quackenbush J., Sedegah M., Shoaibi A., Cummings L.M.,
RA   Florens L., Yates J.R. III, Raine J.D., Sinden R.E., Harris M.A.,
RA   Cunningham D.A., Preiser P.R., Bergman L.W., Vaidya A.B.,
RA   van Lin L.H., Janse C.J., Waters A.P., Smith H.O., White O.R.,
RA   Salzberg S.L., Venter J.C., Fraser C.M., Hoffman S.L., Gardner M.J.,
RA   Carucci D.J.;
RT   "Genome sequence and comparative analysis of the model rodent malaria
RT   parasite Plasmodium yoelii yoelii.";
RL   Nature 419:512-519(2002).
CC   -!- CAUTION: The sequence shown here is derived from an
CC       EMBL/GenBank/DDBJ whole genome shotgun (WGS) entry which is
CC       preliminary data.
DR   EMBL; AABL01002172; EAA18637.1; -; Genomic_DNA.
DR   InterPro; IPR011591; Botulinum.
DR   ProDom; PD001963; Botulinum; 2.
KW   Hypothetical protein.
SQ   SEQUENCE   2649 AA;  309810 MW;  E7207F344643AC24 CRC64;

  Query Match             6.8%;  Score 227;  DB 2;  Length 2649;
  Best Local Similarity   19.0%;  Pred. No. 0.037;
  Matches  131;  Conservative  110;  Mismatches  188;  Indels  260;  Gaps   28;

Qy         15 DAFRSHTFSE-----QTLMSTDLLANSSDPDFMY----------ELDREM--------N 50
              |: : : | :     : |:: :   |||: |::|          ||:  |         |
Db        140 DSAKIYPFDKINNNVKNLLTNNRNDNSSNEDYIYMSNNLNYNFKEELENNMFNNLNMFKN 199

Qy         51 YQQNPRD----NFLSLEDCKDIENLESFTDVLDNEGALTSNWEQWDTYCEDLTKYTKLTS 106
                ::|  |    :|:  :  |||  |    |    || |:  :  : :   | ||: : ||:
Db        200 LKKNNFDDATKHFMKPKTFKDINGLYYDT----NEKAIDIDLLKMEH--EKLTRMSSLTN 253

Qy        107 CDIWGTKEVDYLGLDDFSSPYQDEEVISKTPTLAQLNSEDSQSVSDSLYYPDSLFSVKQN 166
                :    | |||  ::| : | | |        :::|   : : :           :::|
Db        254 LN---RVECDYLAHNNFPNAYNDNEY-----KKSKVNKRGNNNYT----------RLRKN 295

Qy        167 PLPSSFPGKKITSRAAAPVCSSKTLQAEVPLSDCVQKASKPPSSTQIMVKTNMYHNEKVN 226
                |  :   :|                        : | ||  |   :::|:     |
Db        296 ILDKNIRSDEIN-------------------------NTPLSSDDI--TSHIYNRNVEN 327

Qy        227 FHVECKDYVKKAKVKINPVQQSRPLLSQIHTDAAKENTCYCGAVAKRQEKKGMEPLQGHA 286
                ::          | :|   ::  |    :|   :  :||          : || |:  : :
Db        328 YN---------AGLKNEKIENIR----KIGNKSKGKNTILKKIGVHWENKKEMQNEESY- 373

Qy        287 TPALPFKETQELLLSPLPQEGPGSLAAGESSSLSASTSVSDSSQKKEEHNYSLFVSDNLG 346
                                                   |  :|:||
Db        374 ----------------------------------------DEEEKEEEE---------- 382

Qy        347 EQPTKCSPEEDEEDEEDVDDEDHDEGFGSEHELSENEEEE-EEEEDYEDDKDDDISDTFS 405
                    |:||::|:| |||| ::| | | |  | |||| ||||| |:::::|
Db        383 ------VEEDDEDEEDDEDDEDDEDGEGEEDEEEEEEEEEGEEEEDEEEEEEED------ 430

Qy        406 EPGYENDSVEDLKEVTSISSRKR-----GKRRYFWEYSEQLTPSQQERMLRPSEWNRDTL 460
                  |  |: :||    |:|:     ||:|           :  :::|:       |
Db        431 ------DEEEEEEEVEEYESKKKHTYKSGKKR----------NDNDKLLK--------L 465
```

```
Qy        461 PSNMYQKNGLHHGKYAVKKSRRTD-----VEDLTPNPKKLL----QIGNELRK--LNKVI 509
Db        466 KKNNNMSNDVVNNKTNTKSKSRDDNNTKKIQNIKKNIKSFVGTPKGVSKEIVKNIIKRIV 525

Qy        510 SDLTPVSELPLTARPRSRKEKNKLAFRACRLKKKA-------------QYEANKVKL-W 554
Db        526 EDYDNNYSEDFNSNNMTRNVENEDSENNQKLNKNILEPQNTSENEDLINEHIAKKINFEW 585

Qy        555 GLNTEY----------------------DNLLFVINSIKQEIVNRVQNPR---------- 582
Db        586 WKSTEQIDNKKIDGIKKQKDTSDKINNKDNLIEKENSNKEKIGNRYEIPKNISYYVNMNR 645

Qy        583 ---------DERGPNMGQKLEILIKDTLG 602
Db        646 KKNVEKVIGDKNSHNFGEN-----KDSEG 669


RESULT 14
Q7RBX2_PLAYO
ID   Q7RBX2_PLAYO PRELIMINARY;      PRT;   761 AA.
AC   Q7RBX2;
DT   01-MAR-2004 (TrEMBLrel. 26, Created)
DT   01-MAR-2004 (TrEMBLrel. 26, Last sequence update)
DT   01-MAR-2004 (TrEMBLrel. 26, Last annotation update)
DE   Mature-parasite-infected erythrocyte surface antigen.
GN   Name=PY06014;
OS   Plasmodium yoelii yoelii.
OC   Eukaryota; Alveolata; Apicomplexa; Haemosporida; Plasmodium.
OX   NCBI_TaxID=73239;
RN   [1]
RP   NUCLEOTIDE SEQUENCE.
RC   STRAIN=17XNL;
RX   MEDLINE=22255706; PubMed=12368865; DOI=10.1038/nature01099;
RA   Carlton J.M., Angiuoli S.V., Suh B.B., Kooij T.W., Pertea M.,
RA   Silva J.C., Ermolaeva M.D., Allen J.E., Selengut J.D., Koo H.L.,
RA   Peterson J.D., Pop M., Kosack D.S., Shumway M.F., Bidwell S.L.,
RA   Shallom S.J., van Aken S.E., Riedmuller S.B., Feldblyum T.V.,
RA   Cho J.K., Quackenbush J., Sedegah M., Shoaibi A., Cummings L.M.,
RA   Florens L., Yates J.R. III, Raine J.D., Sinden R.E., Harris M.A.,
RA   Cunningham D.A., Preiser P.R., Bergman L.W., Vaidya A.B.,
RA   van Lin L.H., Janse C.J., Waters A.P., Smith H.O., White O.R.,
RA   Salzberg S.L., Venter J.C., Fraser C.M., Hoffman S.L., Gardner M.J.,
RA   Carucci D.J.;
RT   "Genome sequence and comparative analysis of the model rodent malaria
RT   parasite Plasmodium yoelii yoelii.";
RL   Nature 419:512-519(2002).
CC   -!- CAUTION: The sequence shown here is derived from an
CC       EMBL/GenBank/DDBJ whole genome shotgun (WGS) entry which is
CC       preliminary data.
DR   EMBL; AABL01001988; EAA18150.1; -; Genomic_DNA.
SQ   SEQUENCE   761 AA;  90874 MW;  CEEF787A6EA9D5B3 CRC64;

  Query Match             6.3%;  Score 212;  DB 2;  Length 761;
  Best Local Similarity   19.3%;  Pred. No. 0.046;
  Matches  127;  Conservative  106;  Mismatches  239;  Indels  186;  Gaps   23;

Qy         34 ANSSDPDFMYELDREMNYQQNPRDNFLSLEDCKDIENL----------------ESFTD 76
Db         20 SESEDIEIEEENDSDIN-NEIYDDNLVSDSDDESVNNINKKSKGKKKKKKLSSDEESQED 78

Qy         77 VLD--NEGALTSNWEQWDTY----CEDLTKYTKLTSCDIWGTKEVDYLGL--DDFSSPYQ 128
Db         79 DFDDENEGEDLSDDDETENVENKEDEDFIKVS-------WKKDKKNYYQYESDNSSSDDD 131

Qy        129 DEEVISKTPTLAQLNSEDSQSVSDSLYYPDSLFSVKQNPLPSSFPGKKITSRAAAPVCSS 188
Db        132 EENNDERMKEVIYLNKKEKENLNENDF---DLYNIYMNEKDG--------------ISS 173

Qy        189 KTLQAEVPLSDCVQKASKPPSSTQIMVKTNMYHNEKVNFHVECKDYVKKAKVKINPVQQS 248
Db        174 KNKDNEIGIDEKENTIKK-----LINNMANELKEKKKEINIDEKDKKDIEEMIMNEHQEY 228

Qy        249 RPLLSQIHTDAAK----------------ENTCYCGAVAKRQ----EKKGMEPL---- 282
```

```
Db        229 QIILKELSLNIEKVFNEINENQKLFQFKNINENDVSPSDINKNTLLYLKKKNETMLTYII 288

Qy        283 ------------QGHATPALP--------FKETQEL---------LLSPLPQEGPGSLA 312
Db        289 YITYYVFLKVMNCYTHNHPVLDKLIYINTIISKTNELDNKIKFKIQQLNKLPKR------ 342

Qy        313 AGESSSLSASTSVSDSSQKKEEHNYSLFVSDNLGEQPTKCSPEEDEEDEEDV---DDEDH 369
Db        343 --QLDELEISSS-DDQTHTKKNATGAKKIGKKSADNDDEYEEEEDGEDGEDVEAEEDEDE 399

Qy        370 DEGFGSEHELSENEEEEEEEEDYEDDKDDDISDTFSEPGYENDSVEDLKEVTSISSRKRG 429
Db        400 EEEDEEEEEDDEDDEDDEDEEDDEDEEDEDEEDEDDEEEEEEEEEEEEEIEEDKNNKNK 459

Qy        430 KRRYFWEYSEQLTPSQ-QERMLRPSEWNRDTLPSNMYQKNGLHHGKYAVKKSRRT----D 484
Db        460 KYKVSKSIITEYTDSHIREKMKEEKKKQREKI-----------------KNERSIFLKE 501

Qy        485 VEDLTPNPKKLLQIGNELRKLNKVISDLTPVSELPLTARPRSRKEKNKLAFRACRLKKKA 544
Db        502 IKDMVSNKPEKIEEENYLKKLEEKFTDF----------------DDKILRKKMKMMSKK 544

Qy        545 QYEANK--------------VKLWGLNTEYDNLLF--------VINSIKQEIVNRVQN 580
Db        545 KNRMNNLSNVGMTSNDLLKFVELPEMNNENDNTSFHENKIFRNNINKIKQKNKNKLMN 602


RESULT 15
Q8CE30_MOUSE
ID   Q8CE30_MOUSE PRELIMINARY;      PRT;   707 AA.
AC   Q8CE30;
DT   01-MAR-2003 (TrEMBLrel. 23, Created)
DT   01-MAR-2003 (TrEMBLrel. 23, Last sequence update)
DT   01-MAR-2004 (TrEMBLrel. 26, Last annotation update)
DE   Mus musculus 10 days neonate skin cDNA, RIKEN full-length enriched
DE   library, clone:4732495B18 product:nucleolin, full insert
DE   sequence.
GN   Name=Ncl;
OS   Mus musculus (Mouse).
OC   Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;
OC   Mammalia; Eutheria; Euarchontoglires; Glires; Rodentia; Sciurognathi;
OC   Muroidea; Muridae; Murinae; Mus.
OX   NCBI_TaxID=10090;
RN   [1]
RP   NUCLEOTIDE SEQUENCE.
RC   STRAIN=C57BL/6J; TISSUE=Skin;
RX   MEDLINE=99279253; PubMed=10349636; DOI=10.1016/S0076-6879(99)03004-9;
RA   Carninci P., Hayashizaki Y.;
RT   "High-efficiency full-length cDNA cloning.";
RL   Meth. Enzymol. 303:19-44(1999).
RN   [2]
RP   NUCLEOTIDE SEQUENCE.
RC   STRAIN=C57BL/6J; TISSUE=Skin;
RX   MEDLINE=21085660; PubMed=11217851; DOI=10.1038/35055500;
RA   Kawai J., Shinagawa A., Shibata K., Yoshino M., Itoh M., Ishii Y.,
RA   Arakawa T., Hara A., Fukunishi Y., Konno H., Adachi J., Fukuda S.,
RA   Aizawa K., Izawa M., Nishi K., Kiyosawa H., Kondo S., Yamanaka I.,
RA   Saito T., Okazaki Y., Gojobori T., Bono H., Kasukawa T., Saito R.,
RA   Kadota K., Matsuda H.A., Ashburner M., Batalov S., Casavant T.,
RA   Fleischmann W., Gaasterland T., Gissi C., King B., Kochiwa H.,
RA   Kuehl P., Lewis S., Matsuo Y., Nikaido I., Pesole G., Quackenbush J.,
RA   Schriml L.M., Staubli F., Suzuki R., Tomita M., Wagner L., Washio T.,
RA   Sakai K., Okido T., Furuno M., Aono H., Baldarelli R., Barsh G.,
RA   Blake J., Boffelli D., Bojunga N., Carninci P., de Bonaldo M.F.,
RA   Brownstein M.J., Bult C., Fletcher C., Fujita M., Gariboldi M.,
RA   Gustincich S., Hill D., Hofmann M., Hume D.A., Kamiya M., Lee N.H.,
RA   Lyons P., Marchionni L., Mashima J., Mazzarelli J., Mombaerts P.,
RA   Nordone P., Ring B., Ringwald M., Rodriguez I., Sakamoto N.,
RA   Sasaki H., Sato K., Schoenbach C., Seya T., Shibata Y., Storch K.-F.,
RA   Suzuki H., Toyo-oka K., Wang K.H., Weitz C., Whittaker C., Wilming L.,
RA   Wynshaw-Boris A., Yoshida K., Hasegawa Y., Kawaji H., Kohtsuki S.,
RA   Hayashizaki Y.;
```

```
RT   "Functional annotation of a full-length mouse cDNA collection.";
RL   Nature 409:685-690(2001).
RN   [3]
RP   NUCLEOTIDE SEQUENCE.
RC   STRAIN=C57BL/6J; TISSUE=Skin;
RA   The FANTOM Consortium,
RA   the RIKEN Genome Exploration Research Group Phase I & II Team;
RT   "Analysis of the mouse transcriptome based on functional annotation of
RT   60,770 full-length cDNAs.";
RL   Nature 420:563-573(2002).
RN   [4]
RP   NUCLEOTIDE SEQUENCE.
RC   STRAIN=C57BL/6J; TISSUE=Skin;
RX   MEDLINE=20499374; PubMed=11042159; DOI=10.1101/gr.145100;
RA   Carninci P., Shibata Y., Hayatsu N., Sugahara Y., Shibata K., Itoh M.,
RA   Konno H., Okazaki Y., Muramatsu M., Hayashizaki Y.;
RT   "Normalization and subtraction of cap-trapper-selected cDNAs to
RT   prepare full-length cDNA libraries for rapid discovery of new genes.";
RL   Genome Res. 10:1617-1630(2000).
RN   [5]
RP   NUCLEOTIDE SEQUENCE.
RC   STRAIN=C57BL/6J; TISSUE=Skin;
RX   MEDLINE=20530913; PubMed=11076861; DOI=10.1101/gr.152600;
RA   Shibata K., Itoh M., Aizawa K., Nagaoka S., Sasaki N., Carninci P.,
RA   Konno H., Akiyama J., Nishi K., Kitsunai T., Tashiro H., Itoh M.,
RA   Sumi N., Ishii Y., Nakamura S., Hazama M., Nishine T., Harada A.,
RA   Yamamoto R., Matsumoto H., Sakaguchi S., Ikegami T., Kashiwagi K.,
RA   Fujiwake S., Inoue K., Togawa Y., Izawa M., Ohara E., Watahiki M.,
RA   Yoneda Y., Ishikawa T., Ozawa K., Tanaka T., Matsuura S., Kawai J.,
RA   Okazaki Y., Muramatsu M., Inoue Y., Kira A., Hayashizaki Y.;
RT   "RIKEN integrated sequence analysis (RISA) system-384-format
RT   sequencing pipeline with 384 multicapillary sequencer.";
RL   Genome Res. 10:1757-1771(2000).
RN   [6]
RP   NUCLEOTIDE SEQUENCE.
RC   STRAIN=C57BL/6J; TISSUE=Skin;
RA   Adachi J., Aizawa K., Akimura T., Arakawa T., Bono H., Carninci P.,
RA   Fukuda S., Furuno M., Hanagaki T., Hara A., Hashizume W.,
RA   Hayashida K., Hayatsu N., Hiramoto K., Hiraoka T., Hirozane T.,
RA   Hori F., Imotani K., Ishii Y., Itoh M., Kagawa I., Kasukawa T.,
RA   Katoh H., Kawai J., Kojima Y., Kondo S., Konno H., Kouda M., Koya S.,
RA   Kurihara C., Matsuyama T., Miyazaki A., Murata M., Nakamura M.,
RA   Nishi K., Nomura K., Numazaki R., Ohno M., Ohsato N., Okazaki Y.,
RA   Saito R., Saitoh H., Sakai C., Sakai K., Sakazume N., Sano H.,
RA   Sasaki D., Shibata K., Shinagawa A., Shiraki T., Sogabe Y., Tagami M.,
RA   Tagawa A., Takahashi F., Takaku-Akahira S., Takeda Y., Tanaka T.,
RA   Tomaru A., Toya T., Yasunishi A., Muramatsu M., Hayashizaki Y.;
RL   Submitted (JUL-2001) to the EMBL/GenBank/DDBJ databases.
DR   EMBL; AK029128; BAC26311.1; -; mRNA.
DR   HSSP; P08199; 1FJE.
DR   Ensembl; ENSMUSG00000026234; Mus musculus.
DR   MGI; MGI:97286; Ncl.
DR   GO; GO:0005730; C:nucleolus; TAS.
DR   InterPro; IPR012677; a_b_plait_nuc_bd.
DR   InterPro; IPR000504; RNP1_RNA_bd.
DR   Pfam; PF00076; RRM_1; 4.
DR   SMART; SM00360; RRM; 4.
DR   PROSITE; PS50102; RRM; 4.
SQ   SEQUENCE   707 AA;  76712 MW;  476E3E8217F78CF8 CRC64;

  Query Match             6.3%;  Score 210;  DB 2;  Length 707;
  Best Local Similarity   21.0%;  Pred. No. 0.053;
  Matches  128;  Conservative   78;  Mismatches  180;  Indels  224;  Gaps   28;

Qy        137 PTLAQLNSEDSQSVSDSLYYPDSLFSVKQNPLPSSFPGKKITSRAAAPVCSSKTLQAEVP 196
              |   : :||| :   |     |     ::  :|    ||| |:  |  |  |:| :| ||
Db         21 PKEVEEDSEDEEMSEDE----DDSSGEEEVVIPQK-KGKKATTTPAKKVVVSQTKKAAVP 75

Qy        197 LSDCVQKASKPPSSTQIMVKTNMYHNEKVNFHVECKDYVKKAKVKINPVQQSRPLLSQIH 256
                    :||:  |                          |     ||  | | :
Db         76 TP--AKKAAVTPGK--------------------KAVATPAKKNITPAK---------- 102

Qy        257 TDAAKENTCYCGAVAKRQEKKGMEPLQGHATPALPFKETQELLLSPLP-QEGPGSLAAGE 315
```

```
                        |     |||                   |   | | | ::|  : | |
Db        103 ------------VIPTPGKKG--------------AAQAKALVPTPGKKGAATSAKGA 134

Qy        316 SSSLSASTSVSDSSQKKEEHNYSLFVSDN-----------LGEQPTKCSP----EEDEED 360
               :  :|    ||  : :|: : |    |:              | :| | :|     |||||
Db        135 KNGKNAKKEDSDEDEDEEDEDDSDEDEDDEEEDEFEPPIVKGVKPAKAAPAAPASEDEED 194

Qy        361 EEDVDDEDHDEGFGSEHELSENEEE-----------------------EEEEEDYEDDK 396
              :|| |||: |:    | |  ::|||                       |||:| |:|:
Db        195 DEDEDDEEDDD----EEEEDDSEEEVMEITTAKGKKTPAKVVPMKAKSVAEEEDDEEEDE 250

Qy        397 DDDISDTFSEPGYENDSVEDLKEVTSISSRKRGKRRYFWEYSEQL-TPSQQERMLRPSEW 455
              ||:  |    |   ::|  |: :|    :  || |     | ::|   |  ::: :  ||
Db        251 DDEDEDDEEEDDEDDDEEEEEEEPVKAAPGKRKK-----EMTKQKEAPEAKKQKVEGSE- 304

Qy        456 NRDTLPSNMYQKN-----GLHHGKYAVKK---------------SRR------TDVEDL 488
                | | |::  |       ::  |:|: :                :|:        |||
Db        305 --PTTPFNLFIGNLNPNKSVNELKFAISELFAKNDLAVVDVRTGTNRKFGYVDFESAEDL 362

Qy        489 TPNPKKLLQI------GNELRKLNKVISDLTPVSELPLTARPRSRKEKNKLAFRACRLK- 541
                :| |::      |||: || |                  |: |  |   | |    |
Db        363 ----EKALELTGLKVFGNEI-KLEK---------------PKGRDSKKVRAARTLLAKN 401

Qy        542 ----------KKAQYEANKVKLWGLNTEYDNLLFVINSIKQEIVNRVQNPRDERGPNM-- 589
                       |:   :| :::|   : :   : ::   | |  :  :|  :::|  :
Db        402 LSFNITEDELKEVFEDAMEIRLVSQDGKSKGIAYI--EFKSE-ADAEKNLEEKQGAEIDG 458

Qy        590 ----------GQKLE------------ILIKDTLGLPVAGQTSEFVNQVLEKTA---- 621
                       ||: |                |:   |       | | : :| ||
Db        459 RSVSLYYTGEKGQRQERTGKTSTWSGESKTLVLSNLSY---SATKETLEEVFEKATFIKV 515

Qy        622 ----EGNPTG 627
                   | | |
Db        516 PQNPHGKPKG 525
```

Search completed: February 28, 2006, 08:53:00
Job time : 238.353 secs

GenCore version 5.1.7

OM protein - protein search, using sw model

Run on: February 28, 2006, 09:02:21 ; Search time 186 Seconds
(without alignments)
1509.479 Million cell updates/sec

Title: US-10-717-665A-44
Perfect score: 639
Sequence: 1 MPQPSVSGMDPPFGDAFRSH.....TAEGNPTGGLVGLRIPTSKV 639

Scoring table: OLIGO
Gapop 60.0 , Gapext 60.0

Searched: 2443163 seqs, 439378781 residues

Word size : 0

Total number of hits satisfying chosen parameters: 2443163

Minimum DB seq length: 0
Maximum DB seq length: 2000000000

Post-processing: Listing first 45 summaries

Database : A_Geneseq_21:*
1: geneseqp1980s:*
2: geneseqp1990s:*
3: geneseqp2000s:*
4: geneseqp2001s:*
5: geneseqp2002s:*
6: geneseqp2003as:*
7: geneseqp2003bs:*
8: geneseqp2004s:*
9: geneseqp2005s:*

Pred. No. is the number of results predicted by chance to have a score greater than or equal to the score of the result being printed, and is derived by analysis of the total score distribution.

SUMMARIES

Result No.	Score	% Query Match	Length	DB	ID	Description
1	639	100.0	639	7	ADK65805	Adk65805 Angiogene
2	307	48.0	417	8	ADQ66472	Adq66472 Novel hum
3	200	31.3	256	5	ADK35935	Adk35935 Novel hum
4	164	25.7	168	5	ABP64928	Abp64928 Human pro
5	150	23.5	296	8	ADO20373	Ado20373 Human PRO
6	83	13.0	91	7	ADK65836	Adk65836 Angiogene
7	40	6.3	160	4	ABG29275	Abg29275 Novel hum
8	11	1.7	830	5	ABB04812	Abb04812 LDL recep
9	11	1.7	1100	4	AAG84930	Aag84930 Shrimp wh
10	11	1.7	1174	4	AAG85039	Aag85039 Shrimp wh
11	11	1.7	1483	7	ADF69140	Adf69140 Human MP5
12	11	1.7	1483	8	ADQ20910	Adq20910 Human sof
13	11	1.7	1483	9	ADX05853	Adx05853 Cyclin-de
14	11	1.7	1483	9	ADX05855	Adx05855 Cyclin-de
15	11	1.7	1527	2	AAW81172	Aaw81172 Human BAZ
16	11	1.7	1531	2	AAW81173	Aaw81173 Human BAZ
17	10	1.6	158	8	ADH45438	Adh45438 Human mol
18	10	1.6	179	3	AAG36936	Aag36936 Arabidops
19	10	1.6	236	8	ADJ46862	Adj46862 Human BIC
20	10	1.6	236	8	ADP55980	Adp55980 Human PRO
21	10	1.6	255	3	AAY66717	Aay66717 Membrane-
22	10	1.6	255	4	AAU29131	Aau29131 Human PRO
23	10	1.6	255	4	AAB65240	Aab65240 Human PRO
24	10	1.6	255	6	ABU58507	Abu58507 Human PRO
25	10	1.6	255	6	ABU88055	Abu88055 Novel hum
26	10	1.6	255	6	ABU84370	Abu84370 Human sec
27	10	1.6	255	6	ABR66244	Abr66244 Human sec
28	10	1.6	255	6	ABR65634	Abr65634 Human sec
29	10	1.6	255	6	ABU99574	Abu99574 Human sec
30	10	1.6	255	6	ABU58055	Abu58055 Human PRO
31	10	1.6	255	6	ABU59133	Abu59133 Novel hum
32	10	1.6	255	6	ABU82645	Abu82645 Human sec
33	10	1.6	255	6	ABU82813	Abu82813 Human PRO
34	10	1.6	255	6	ABU89934	Abu89934 Novel hum
35	10	1.6	255	6	ABR68183	Abr68183 Human sec
36	10	1.6	255	6	ABU60564	Abu60564 Human sec
37	10	1.6	255	6	ABU96236	Abu96236 Novel hum
38	10	1.6	255	6	ABU92667	Abu92667 Human sec
39	10	1.6	255	6	ABO08744	Abo08744 Human sec
40	10	1.6	255	6	ABO02796	Abo02796 Human sec
41	10	1.6	255	6	ABR74950	Abr74950 Human sec
42	10	1.6	255	6	ABR94712	Abr94712 Human sec
43	10	1.6	255	6	ABU13946	Abu13946 Human PRO
44	10	1.6	255	6	ABU85685	Abu85685 Human PRO
45	10	1.6	255	6	ABU98845	Abu98845 Novel hum

ALIGNMENTS

RESULT 1
ADK65805
ID ADK65805 standard; protein; 639 AA.
XX
AC ADK65805;
XX
DT 06-MAY-2004 (first entry)
XX
DE Angiogenesis-differentially expressed protein ANH0757.
XX
KW cytostatic; cardiant; vasotropic; antiarteriosclerotic;
KW angiogenesis inhibitor; angiogenesis stimulator; angiogenic index;
KW gene expression; cancer; coronary artery disease; myocardial ischemia;
KW coronary arteriosclerosis; forensic medicine.
XX
OS Homo sapiens.
XX
PN WO2003066831-A2.
XX
PD 14-AUG-2003.
XX
PF 07-FEB-2003; 2003WO-US003848.
XX
PR 07-FEB-2002; 2002US-00067482.
PR 10-JUN-2002; 2002US-00164595.
PR 16-AUG-2002; 2002US-0403649P.
PR 03-JAN-2003; 2003US-0437746P.
XX
PA (ORIG-) ORIGENE TECHNOLOGIES INC.
XX
PI Sun Z, Li X, Kovacs KF, Fan W, Jay G;
XX
DR WPI; 2003-731502/69.
DR N-PSDB; ADK65804.
XX
PT Determining the angiogenic index of a tissue or cell sample using
PT expression levels of differentially expressed genes, useful for
PT diagnosing or treating cancer, coronary artery disease, myocardial
PT ischemia and/or arteriosclerosis.
XX
PS Claim 23; SEQ ID NO 44; 296pp; English.
XX
CC The invention relates to a method of determining the angiogenic index of
CC a tissue or cell sample comprising assessing, in a sample, the expression
CC levels of one or more differentially-expressed gene from any of 34 DNA
CC sequences, given in the specification, where the levels are indicative of
CC the angiogenic index. The methods and compositions of the present

```
CC   invention are useful for diagnosing, preventing and/or treating cancer,
CC   coronary artery disease, myocardial ischemia or coronary
CC   arteriosclerosis. They can also be used in research, drug discovery and
CC   forensic medicine involving angiogenesis. This sequence corresponds to
CC   one of the differentially expressed proteins of the invention.
XX
SQ   Sequence 639 AA;

  Query Match             100.0%;  Score 639;  DB 7;  Length 639;
  Best Local Similarity   100.0%;  Pred. No. 0;
  Matches  639;  Conservative    0;  Mismatches    0;  Indels    0;  Gaps    0;

Qy          1 MPQPSVSGMDPPFGDAFRSHTFSEQTLMSTDLLANSSDPDFMYELDREMNYQQNPRDNFL 60
              ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db          1 MPQPSVSGMDPPFGDAFRSHTFSEQTLMSTDLLANSSDPDFMYELDREMNYQQNPRDNFL 60

Qy         61 SLEDCKDIENLESFTDVLDNEGALTSNWEQWDTYCEDLTKYTKLTSCDIWGTKEVDYLGL 120
              ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db         61 SLEDCKDIENLESFTDVLDNEGALTSNWEQWDTYCEDLTKYTKLTSCDIWGTKEVDYLGL 120

Qy        121 DDFSSPYQDEEVISKTPTLAQLNSEDSQSVSDSLYYPDSLFSVKQNPLPSSFPGKKITSR 180
              ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db        121 DDFSSPYQDEEVISKTPTLAQLNSEDSQSVSDSLYYPDSLFSVKQNPLPSSFPGKKITSR 180

Qy        181 AAAPVCSSKTLQAEVPLSDCVQKASKPPSSTQIMVKTNMYHNEKVNFHVECKDYVKKAKV 240
              ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db        181 AAAPVCSSKTLQAEVPLSDCVQKASKPPSSTQIMVKTNMYHNEKVNFHVECKDYVKKAKV 240

Qy        241 KINPVQQSRPLLSQIHTDAAKENTCYCGAVAKRQEKKGMEPLQGHATPALPFKETQELLL 300
              ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db        241 KINPVQQSRPLLSQIHTDAAKENTCYCGAVAKRQEKKGMEPLQGHATPALPFKETQELLL 300

Qy        301 SPLPQEGPGSLAAGESSSLSASTSVSDSSQKKEEHNYSLFVSDNLGEQPTKCSPEEDEED 360
              ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db        301 SPLPQEGPGSLAAGESSSLSASTSVSDSSQKKEEHNYSLFVSDNLGEQPTKCSPEEDEED 360

Qy        361 EEDVDDEDHDEGFGSEHELSENEEEEEEEEDYEDDKDDDISDTFSEPGYENDSVEDLKEV 420
              ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db        361 EEDVDDEDHDEGFGSEHELSENEEEEEEEEDYEDDKDDDISDTFSEPGYENDSVEDLKEV 420

Qy        421 TSISSRKRGKRRYFWEYSEQLTPSQQERMLRPSEWNRDTLPSNMYQKNGLHHGKYAVKKS 480
              ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db        421 TSISSRKRGKRRYFWEYSEQLTPSQQERMLRPSEWNRDTLPSNMYQKNGLHHGKYAVKKS 480

Qy        481 RRTDVEDLTPNPKKLLQIGNELRKLNKVISDLTPVSELPLTARPRSRKEKNKLAFRACRL 540
              ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db        481 RRTDVEDLTPNPKKLLQIGNELRKLNKVISDLTPVSELPLTARPRSRKEKNKLAFRACRL 540

Qy        541 KKKAQYEANKVKLWGLNTEYDNLLFVINSIKQEIVNRVQNPRDERGPNMGQKLEILIKDT 600
              ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db        541 KKKAQYEANKVKLWGLNTEYDNLLFVINSIKQEIVNRVQNPRDERGPNMGQKLEILIKDT 600

Qy        601 LGLPVAGQTSEFVNQVLEKTAEGNPTGGLVGLRIPTSKV 639
              ||||||||||||||||||||||||||||||||||||||
Db        601 LGLPVAGQTSEFVNQVLEKTAEGNPTGGLVGLRIPTSKV 639


RESULT 2
ADQ66472
ID   ADQ66472 standard; protein; 417 AA.
XX
AC   ADQ66472;
XX
DT   07-OCT-2004  (first entry)
XX
DE   Novel human protein sequence #1445.
XX
KW   osteopathic; neuroprotective; nootropic; antiparkinsonian; cytostatic;
KW   gene therapy; diagnostic marker; morbid state; osteoporosis;
KW   neurological disease; Alzheimer's disease; Parkinson's disease; dementia;
KW   cancer.
XX
QS   Homo sapiens.
XX
PN   EP1440981-A2.
XX
PD   28-JUL-2004.
XX
PF   21-JAN-2004; 2004EP-00001196.
XX
PR   21-JAN-2003; 2003JP-00102206.
PR   09-MAY-2003; 2003JP-00131392.
XX
PA   (REAS-) RES ASSOC BIOTECHNOLOGY.
XX
PI   Isogai T,  Sugiyama T,  Otsuki T,  Wakamatsu A,  Sato H,  Ishii S;
PI   Yamamoto J,  Isono Y,  Nagai K,  Irie R;
XX
DR   WPI; 2004-535376/52.
DR   N-PSDB; ADQ64284.
XX
PT   Novel 2495 cDNA, useful for treating osteoporosis, neurological diseases,
PT   Alzheimer's diseases, Parkinson's diseases, dementia and various cancers.
XX
PS   Claim 1; SEQ ID NO 3633; 2449pp; English.
XX
CC   The invention relates to 2495 novel polynucleotides (I) and their encoded
CC   polypeptides, sequences hybridizing to these nucleotides, sequences
CC   encoding partial polypeptides and sequences having 70% or 90% identity to
CC   the nucleotide and protein sequences. The nucleotides and polypeptides
CC   are useful as diagnostic markers or therapeutic target for the diseases
CC   or morbid states. They are also useful for treating osteoporosis,
CC   neurological diseases, Alzheimer's diseases, Parkinson's diseases,
CC   dementia and various cancers. This sequence corresponds to a protein
CC   sequence of the invention.
XX
SQ   Sequence 417 AA;

  Query Match             48.0%;  Score 307;  DB 8;  Length 417;
  Best Local Similarity   99.8%;  Pred. No. 4e-290;
  Matches  407;  Conservative    0;  Mismatches    1;  Indels    0;  Gaps    0;

Qy          1 MPQPSVSGMDPPFGDAFRSHTFSEQTLMSTDLLANSSDPDFMYELDREMNYQQNPRDNFL 60
              ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db          1 MPQPSVSGMDPPFGDAFRSHTFSEQTLMSTDLLANSSDPDFMYELDREMNYQQNPRDNFL 60

Qy         61 SLEDCKDIENLESFTDVLDNEGALTSNWEQWDTYCEDLTKYTKLTSCDIWGTKEVDYLGL 120
              ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db         61 SLEDCKDIENLESFTDVLDNEGALTSNWEQWDTYCEDLTKYTKLTSCDIWGTKEVDYLGL 120

Qy        121 DDFSSPYQDEEVISKTPTLAQLNSEDSQSVSDSLYYPDSLFSVKQNPLPSSFPGKKITSR 180
              ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db        121 DDFSSPYQDEEVISKTPTLAQLNSEDSQSVSDSLYYPDSLFSVKQNPLPSSFPGKKITSR 180

Qy        181 AAAPVCSSKTLQAEVPLSDCVQKASKPPSSTQIMVKTNMYHNEKVNFHVECKDYVKKAKV 240
              |||||||||||||||||||||||||| ||||||||||||||||||||||||||||||||||
Db        181 AAAPVCSSKTLQAEVPLSDCVQKASKPTSSTQIMVKTNMYHNEKVNFHVECKDYVKKAKV 240

Qy        241 KINPVQQSRPLLSQIHTDAAKENTCYCGAVAKRQEKKGMEPLQGHATPALPFKETQELLL 300
              ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db        241 KINPVQQSRPLLSQIHTDAAKENTCYCGAVAKRQEKKGMEPLQGHATPALPFKETQELLL 300

Qy        301 SPLPQEGPGSLAAGESSSLSASTSVSDSSQKKEEHNYSLFVSDNLGEQPTKCSPEEDEED 360
              ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db        301 SPLPQEGPGSLAAGESSSLSASTSVSDSSQKKEEHNYSLFVSDNLGEQPTKCSPEEDEED 360

Qy        361 EEDVDDEDHDEGFGSEHELSENEEEEEEEEDYEDDKDDDISDTFSEPG 408
              ||||||||||||||||||||||||||||||||||||||||||||||||
Db        361 EEDVDDEDHDEGFGSEHELSENEEEEEEEEDYEDDKDDDISDTFSEPG 408


RESULT 3
ADK35935
ID   ADK35935 standard; protein; 256 AA.
XX
AC   ADK35935;
```

```
XX
DT   06-MAY-2004  (first entry)
XX
DE   Novel human polypeptide SeqID8017.
XX
KW   antiarthritic; antiparkinsonian; neuroprotective; nootropic;
KW   immunosuppressive; cytostatic; antipsoriatic; antiinflammatory;
KW   antibacterial; antiviral; antifungal; antiparasitic; gene therapy;
KW   arthritis; Parkinson's; Alzheimer's; autoimmune disease; cancer;
KW   psoriasis; inflammatory bowel disease; infection; bacteria; virus;
KW   fungus; parasite; human.
XX
OS   Homo sapiens.
XX
FH   Key             Location/Qualifiers
FT   Misc-difference 1. .256
FT                   /label= OTHER
FT                   /note= "OTHER= All Xaa's in this sequence are unknown
FT                   amino acids or the site of a stop codon within the DNA
FT                   sequence"
XX
PN   WO200216439-A2.
XX
PD   28-FEB-2002.
XX
PF   05-MAR-2001; 2001WO-US004941.
XX
PR   07-MAR-2000; 2000US-00519705.
PR   19-MAY-2000; 2000US-00574454.
XX
PA   (HYSE-) HYSEQ INC.
XX
PI   Tang YT,  Liu C,  Drmanac RT;
XX
DR   WPI; 2002-280918/32.
XX
PT   Isolated polynucleotide encoding bone marrow derived polypeptides useful
PT   for treating, e.g., Parkinson's, Alzheimer's, cancer, arthritis, Crohn's
PT   disease, and inflammatory bowel disease.
XX
PS   Claim 20; SEQ ID NO 8017; 504pp; English.
XX
CC   This invention relates to a novel isolated polynucleotide comprising a
CC   nucleotide sequence selected from one of 1680 sequences, a mature protein
CC   coding portion of them, an active domain of them and their complementary
CC   sequences. The invention may be useful for the production of compounds
CC   with an antiarthritic, antiparkinsonian, neuroprotective, nootropic,
CC   immunosuppressive, cytostatic, antipsoriatic, antiinflammatory,
CC   antibacterial, antiviral, antifungal or antiparasitic activity. In
CC   addition, the disclosed sequences may be useful for gene therapy. The
CC   polypeptides or their antibodies are useful for treating many diseases
CC   such as arthritis, Parkinson's, Alzheimer's, autoimmune diseases, cancer,
CC   psoriasis, inflammatory bowel disease and infections caused by bacteria,
CC   viruses, fungi or parasites. The present sequence is that of a human
CC   polypeptide of the invention.
XX
SQ   Sequence 256 AA;

  Query Match             31.3%;  Score 200;  DB 5;  Length 256;
  Best Local Similarity   100.0%;  Pred. No. 6e-186;
  Matches  200;  Conservative    0;  Mismatches    0;  Indels    0;  Gaps    0;

Qy        209 SSTQIMVKTNMYHNEKVNFHVECKDYVKKAKVKINPVQQSRPLLSQIHTDAAKENTCYCG 268
              ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db         51 SSTQIMVKTNMYHNEKVNFHVECKDYVKKAKVKINPVQQSRPLLSQIHTDAAKENTCYCG 110

Qy        269 AVAKRQEKKGMEPLQGHATPALPFKETQELLLSPLPQEGPGSLAAGESSSLSASTSVSDS 328
              ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db        111 AVAKRQEKKGMEPLQGHATPALPFKETQELLLSPLPQEGPGSLAAGESSSLSASTSVSDS 170

Qy        329 SQKKEEHNYSLFVSDNLGEQPTKCSPEEDEEDEEDVDDEDHDEGFGSEHELSENEEEEEE 388
              ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db        171 SQKKEEHNYSLFVSDNLGEQPTKCSPEEDEEDEEDVDDEDHDEGFGSEHELSENEEEEEE 230

Qy        389 EEDYEDDKDDDISDTFSEPG 408
              ||||||||||||||||||||
Db        231 EEDYEDDKDDDISDTFSEPG 250


RESULT 4
ABP64928
ID   ABP64928 standard; protein; 168 AA.
XX
AC   ABP64928;
XX
DT   25-FEB-2003  (first entry)
XX
DE   Human protein SEQ ID 588.
XX
KW   Human; expressed sequence tag; EST; haematopoietic disorder;
KW   central nervous system disease; viral infection;
KW   peripheral nervous system disease; non-healing wound; infectious disease;
KW   immune deficiency; immune disorder; bacterial infection; allergy; cancer;
KW   fungal infection; autoimmune disorder; coagulation disorder; nootropic;
KW   antiallergic; antiinflammatory; immunosuppressive; neuroprotective;
KW   cytostatic; haemostatic; virucide; antibacterial; fungicide;
KW   immunostimulant; cerebroprotective.
XX
OS   Homo sapiens.
XX
PN   WO200259260-A2.
XX
PD   01-AUG-2002.
XX
PF   16-NOV-2001; 2001WO-US042950.
XX
PR   17-NOV-2000; 2000US-00714936.
XX
PA   (HYSE-) HYSEQ INC.
XX
PI   Tang YT,  Goodrich RW,  Liu C,  Zhou P,  Asundi V,  Zhang J,  Zhao QA;
PI   Ren F,  Xue AJ,  Yang Y,  Wehrman T,  Drmanac RT;
XX
DR   WPI; 2002-590824/63.
DR   N-PSDB; ABQ99514.
XX
PT   New isolated polynucleotide, useful in research, diagnostic or
PT   therapeutic methods, e.g. preventing or treating disorders involving
PT   aberrant protein expression or biological activity.
XX
PS   Claim 20; SEQ ID NO 588; 394pp; English.
XX
CC   The present invention relates to novel human coding sequences (ABQ99268-
CC   ABQ99608) and proteins (ABP64682-ABP65022). The sequences are useful in
CC   therapeutic, diagnostic and research methods. The polynucleotides may be
CC   used in the field of molecular biology as hybridisation probes, primers
CC   for PCR, for chromosome and gene mapping, for the recombinant production
CC   of protein, or in generation of anti-sense DNA or RNA. The
CC   polynucleotides are useful in diagnostics as expressed sequence tags
CC   (ESTs) for identifying expressed genes or for physical mapping of the
CC   human genome. The proteins may be used as molecular weight markers, or as
CC   nutritional sources or supplements. The proteins may be used to maintain
CC   and expand cell population in a totipotential or pluripotential state
CC   useful for re-engineering damaged or diseased tissues, transplantation,
CC   manufacture of bio-pharmaceuticals or the development of bio-sensors. The
CC   polynucleotides and proteins are useful for preventing, treating or
CC   ameliorating disorders involving aberrant protein expression or
CC   biological activity, e.g. haematopoietic disorders, central/peripheral
CC   nervous system diseases, mechanical and traumatic disorders, non-healing
CC   wounds, immune deficiencies and disorders, infectious diseases caused by
CC   viral, bacterial or fungal infection, autoimmune disorders, allergic
CC   reactions and conditions, coagulation disorders, or cancer. The
CC   polynucleotide sequences of the invention were assembled from ESTs
CC   isolated mainly by sequencing by hybridisation, and in some cases,
CC   sequences obtained from one or more public databases. Note: The sequence
CC   data for this patent did not form part of the printed specification, but
```

```
CC   was obtained in electronic format directly from WIPO at
CC   ftp.wipo.int/pub/published_pct_sequences
XX
SQ   Sequence 168 AA;

  Query Match             25.7%;  Score 164;  DB 5;  Length 168;
  Best Local Similarity   100.0%;  Pred. No. 5.4e-151;
  Matches  164;  Conservative    0;  Mismatches    0;  Indels    0;  Gaps    0;

Qy        214 MVKTNMYHNEKVNFHVECKDYVKKAKVKINPVQQSRPLLSQIHTDAAKENTCYCGAVAKR 273
              ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db          1 MVKTNMYHNEKVNFHVECKDYVKKAKVKINPVQQSRPLLSQIHTDAAKENTCYCGAVAKR 60

Qy        274 QEKKGMEPLQGHATPALPFKETQELLLSPLPQEGPGSLAAGESSSLSASTSVSDSSQKKE 333
              ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db         61 QEKKGMEPLQGHATPALPFKETQELLLSPLPQEGPGSLAAGESSSLSASTSVSDSSQKKE 120

Qy        334 EHNYSLFVSDNLGEQPTKCSPEEDEEDEEDVDDEDHDEGFGSEH 377
              |||||||||||||||||||||||||||||||||||||||||||||
Db        121 EHNYSLFVSDNLGEQPTKCSPEEDEEDEEDVDDEDHDEGFGSEH 164


RESULT 5
ADO20373
ID   ADO20373 standard; protein; 296 AA.
XX
AC   ADO20373;
XX
DT   12-AUG-2004  (first entry)
XX
DE   Human PRO polypeptide #633.
XX
KW   Human; PRO; immune related disorder; systemic lupus erythematosus;
KW   rheumatoid arthritis; osteoarthritis; juvenile chronic arthritis;
KW   systemic sclerosis; Sjogren's syndrome; vasculitis; sarcoidosis;
KW   autoimmune haemolytic anaemia; autoimmune thrombocytopenia; thyroiditis;
KW   diabetes mellitus; renal disease; demyelinating disease;
KW   central nervous system; peripheral nervous system;
KW   demyelinating polyneuropathy; Guillain-Barre syndrome;
KW   chronic inflammatory demyelinating polyneuropathy.
XX
OS   Homo sapiens.
XX
PN   WO2004043361-A2.
XX
PD   27-MAY-2004.
XX
PF   06-NOV-2003; 2003WO-US035268.
XX
PR   08-NOV-2002; 2002US-0425235P.
XX
PA   (GETH ) GENENTECH INC.
XX
PI   Fong S,  Dennis K,  Clark H,  Chiu H,  Schoenfeld J,  Williams PM;
PI   Wood WI,  Wu TD;
XX
DR   WPI; 2004-420067/39.
DR   N-PSDB; ADO20372.
XX
PT   Novel PRO polypeptide e.g., PRO69614, PRO71106, or PRO86388 useful for
PT   treating an immune related disorder such as systemic lupus erythematosus,
PT   rheumatoid arthritis, osteoarthritis, juvenile chronic arthritis or
PT   spondyloarthropathy.
XX
PS   Claim 7; SEQ ID NO 1266; 1731pp; English.
XX
CC   The invention relates to human PRO polypeptides and the polynucleotides
CC   encoding them. The polypeptides and polynucleotides are useful for
CC   treating and diagnosing immune related disorders in mammals. The immune
CC   related disorders include systemic lupus erythematosus, rheumatoid
CC   arthritis, osteoarthritis, juvenile chronic arthritis, systemic
CC   sclerosis, Sjogren's syndrome, vasculitis, sarcoidosis, autoimmune
CC   haemolytic anaemia, autoimmune thrombocytopenia, thyroiditis, diabetes
CC   mellitus, immune-mediated renal disease, demyelinating diseases of the
CC   central or peripheral nervous system, demyelinating polyneuropathy,
CC   Guillain-Barre syndrome and chronic inflammatory demyelinating
CC   polyneuropathy. This sequence represents a human PRO polypeptide of the
CC   invention.
XX
SQ   Sequence 296 AA;

  Query Match             23.5%;  Score 150;  DB 8;  Length 296;
  Best Local Similarity   99.6%;  Pred. No. 4e-137;
  Matches  250;  Conservative    0;  Mismatches    1;  Indels    0;  Gaps    0;

Qy        389 EEDYEDDKDDDISDTFSEPGYENDSVEDLKEVTSISSRKRGKRRYFWEYSEQLTPSQQER 448
              ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db         46 EEDYEDDKDDDISDTFSEPGYENDSVEDLKEVTSISSRKRGKRRYFWEYSEQLTPSQQER 105

Qy        449 MLRPSEWNRDTLPSNMYQKNGLHHGKYAVKKSRRTDVEDLTPNPKKLLQIGNELRKLNKV 508
              ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db        106 MLRPSEWNRDTLPSNMYQKNGLHHGKYAVKKSRRTDVEDLTPNPKKLLQIGNELRKLNKV 165

Qy        509 ISDLTPVSELPLTARPRSRKEKNKLAFRACRLKKKAQYEANKVKLWGLNTEYDNLLFVIN 568
              |||||||||||||||||||||||| ||||||||||||||||||||||||||||||||||
Db        166 ISDLTPVSELPLTARPRSRKEKNKLASRACRLKKKAQYEANKVKLWGLNTEYDNLLFVIN 225

Qy        569 SIKQEIVNRVQNPRDERGPNMGQKLEILIKDTLGLPVAGQTSEFVNQVLEKTAEGNPTGG 628
              ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db        226 SIKQEIVNRVQNPRDERGPNMGQKLEILIKDTLGLPVAGQTSEFVNQVLEKTAEGNPTGG 285

Qy        629 LVGLRIPTSKV 639
              ||||||||||| 
Db        286 LVGLRIPTSKV 296


RESULT 6
ADK65836
ID   ADK65836 standard; protein; 91 AA.
XX
AC   ADK65836;
XX
DT   06-MAY-2004  (first entry)
XX
DE   Angiogenesis-differentially expressed protein #50.
XX
KW   cytostatic; cardiant; vasotropic; antiarteriosclerotic;
KW   angiogenesis inhibitor; angiogenesis stimulator; angiogenic index;
KW   gene expression; cancer; coronary artery disease; myocardial ischemia;
KW   coronary arteriosclerosis; forensic medicine.
XX
OS   Homo sapiens.
XX
PN   WO2003066831-A2.
XX
PD   14-AUG-2003.
XX
PF   07-FEB-2003; 2003WO-US003848.
XX
PR   07-FEB-2002; 2002US-00067482.
PR   10-JUN-2002; 2002US-00164595.
PR   16-AUG-2002; 2002US-0403649P.
PR   03-JAN-2003; 2003US-0437746P.
XX
PA   (ORIG-) ORIGENE TECHNOLOGIES INC.
XX
PI   Sun Z,  Li X,  Kovacs KF,  Fan W,  Jay G;
XX
DR   WPI; 2003-731502/69.
XX
PT   Determining the angiogenic index of a tissue or cell sample using
PT   expression levels of differentially expressed genes, useful for
PT   diagnosing or treating cancer, coronary artery disease, myocardial
PT   ischemia and/or arteriosclerosis.
XX
PS   Disclosure; SEQ ID NO 75; 296pp; English.
```

XX
CC The invention relates to a method of determining the angiogenic index of
CC a tissue or cell sample comprising assessing, in a sample, the expression
CC levels of one or more differentially-expressed gene from any of 34 DNA
CC sequences, given in the specification, where the levels are indicative of
CC the angiogenic index. The methods and compositions of the present
CC invention are useful for diagnosing, preventing and/or treating cancer,
CC coronary artery disease, myocardial ischemia or coronary
CC arteriosclerosis. They can also be used in research, drug discovery and
CC forensic medicine involving angiogenesis. This sequence corresponds to
CC one of the differentially expressed proteins of the invention.
XX
SQ Sequence 91 AA;

Query Match 13.0%; Score 83; DB 7; Length 91;
Best Local Similarity 100.0%; Pred. No. 3.1e-72;
Matches 83; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

Qy 449 MLRPSEWNRDTLPSNMYQKNGLHHGKYAVKKSRRTDVEDLTPNPKKLLQIGNELRKLNKV 508
||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db 1 MLRPSEWNRDTLPSNMYQKNGLHHGKYAVKKSRRTDVEDLTPNPKKLLQIGNELRKLNKV 60

Qy 509 ISDLTPVSELPLTARPRSRKEKN 531
|||||||||||||||||||||||
Db 61 ISDLTPVSELPLTARPRSRKEKN 83

RESULT 7
ABG29275
ID ABG29275 standard; protein; 160 AA.
XX
AC ABG29275;
XX
DT 18-FEB-2002 (first entry)
XX
DE Novel human diagnostic protein #29266.
XX
KW Human; chromosome mapping; gene mapping; gene therapy; forensic;
KW food supplement; medical imaging; diagnostic; genetic disorder.
XX
OS Homo sapiens.
XX
PN WO200175067-A2.
XX
PD 11-OCT-2001.
XX
PF 30-MAR-2001; 2001WO-US008631.
XX
PR 31-MAR-2000; 2000US-00540217.
PR 23-AUG-2000; 2000US-00649167.
XX
PA (HYSE-) HYSEQ INC.
XX
PI Drmanac RT, Liu C, Tang YT;
XX
DR WPI; 2001-639362/73.
DR N-PSDB; AAS93462.
XX
PT New isolated polynucleotide and encoded polypeptides, useful in
PT diagnostics, forensics, gene mapping, identification of mutations
PT responsible for genetic disorders or other traits and to assess
PT biodiversity.
XX
PS Claim 20; SEQ ID NO 59634; 103pp; English.
XX
CC The invention relates to isolated polynucleotide (I) and polypeptide (II)
CC sequences. (I) is useful as hybridisation probes, polymerase chain
CC reaction (PCR) primers, oligomers, and for chromosome and gene mapping,
CC and in recombinant production of (II). The polynucleotides are also used
CC in diagnostics as expressed sequence tags for identifying expressed
CC genes. (I) is useful in gene therapy techniques to restore normal
CC activity of (II) or to treat disease states involving (II). (II) is
CC useful for generating antibodies against it, detecting or quantitating a
CC polypeptide in tissue, as molecular weight markers and as a food
CC supplement. (II) and its binding partners are useful in medical imaging
CC of sites expressing (II). (I) and (II) are useful for treating disorders
CC involving aberrant protein expression or biological activity. The
CC polypeptide and polynucleotide sequences have applications in
CC diagnostics, forensics, gene mapping, identification of mutations
CC responsible for genetic disorders or other traits to assess biodiversity
CC and to produce other types of data and products dependent on DNA and
CC amino acid sequences. ABG00010-ABG30377 represent novel human diagnostic
CC amino acid sequences of the invention. Note: The sequence data for this
CC patent did not appear in the printed specification, but was obtained in
CC electronic format directly from WIPO at
CC ftp.wipo.int/pub/published_pct_sequences
XX
SQ Sequence 160 AA;

Query Match 6.3%; Score 40; DB 4; Length 160;
Best Local Similarity 100.0%; Pred. No. 4.4e-30;
Matches 40; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

Qy 562 NLLFVINSIKQEIVNRVQNPRDERGPNMGQKLEILIKDTL 601
||||||||||||||||||||||||||||||||||||||||
Db 29 NLLFVINSIKQEIVNRVQNPRDERGPNMGQKLEILIKDTL 68

RESULT 8
ABB04812
ID ABB04812 standard; protein; 830 AA.
XX
AC ABB04812;
XX
DT 13-MAR-2002 (first entry)
XX
DE LDL receptor binding protein JIP-2 SEQ ID NO:35.
XX
KW Low density lipoprotein receptor binding protein; signal transduction;
KW LDL receptor binding protein; LDL receptor signalling pathway.
XX
OS Synthetic.
XX
PN WO200184159-A2.
XX
PD 08-NOV-2001.
XX
PF 24-APR-2001; 2001WO-US013214.
XX
PR 01-MAY-2000; 2000US-00562737.
XX
PA (TEXA) UNIV TEXAS SYSTEM.
XX
PI Herz J, Gotthardt M;
XX
DR WPI; 2002-082855/11.
XX
PT Detecting stress that alters interaction of LDL receptor binding
PT polypeptide with LDL receptor interaction domain, comprises detecting
PT difference in stress-biased and unbiased interaction of peptide and
PT domain in a system.
XX
PS Disclosure; Page 78-79; 200pp; English.
XX
CC The present invention describes a method for detecting a stress that
CC alters a functional interaction of a low density lipoprotein (LDL)
CC receptor binding protein (I) with an LDL receptor interaction domain
CC (II). The method involves introducing a predetermined stress into a
CC system which provides a stress-biased physical interaction of (I) with
CC (II), where in the absence of the stress, the system provides an unbiased
CC interaction of (I) and (II), and detecting the stress-biased interaction
CC of (I) and (II), where a difference between BI and UI indicates that the
CC stress alters the interaction of (I) and (II). (I) is selected from
CC SEMCAP-1, JIP-1, PSD-95, JIP-2, Talin, OMP25, CAPON, PIP4,5 Kinase, Na
CC channel brain 3, Mint1, ICAP-1 and APC subunit 10. The method is useful
CC for detecting a stress that alters functional interaction of LDL receptor

```
CC   binding polypeptide with LDL receptor interaction domain. The method is
CC   useful for detecting and modulating signal transduction through LDL
CC   receptors. ABB04778 to ABB04909 represent LDL receptor binding proteins
CC   which are used in the exemplification of the present invention
XX
SQ   Sequence 830 AA;

  Query Match             1.7%;  Score 11;  DB 5;  Length 830;
  Best Local Similarity   100.0%;  Pred. No. 0.36;
  Matches   11;  Conservative    0;  Mismatches    0;  Indels    0;  Gaps    0;

Qy        381 ENEEEEEEEED 391
              ||||||||||
Db         94 ENEEEEEEEED 104


RESULT 9
AAG84930
ID   AAG84930 standard; protein; 1100 AA.
XX
AC   AAG84930;
XX
DT   06-AUG-2003  (revised)
DT   11-SEP-2001  (first entry)
XX
DE   Shrimp white spot Bacilliform virus (WSBV) protein 21.
XX
KW   Shrimp white spot Bacilliform virus; WSBV; diagnosis; viral infection;
KW   antiviral agent; gene expression; antisense construct;
KW   transgenic viral resistant shrimp.
XX
OS   White spot syndrome virus.
XX
PN   WO200138351-A2.
XX
PD   31-MAY-2001.
XX
PF   08-NOV-2000; 2000WO-US028888.
XX
PR   24-NOV-1999;   99CN-00124717.
XX
PA   (PENY-) PE CORP NY.
PA   (THIR-) THIRD INST OCEANOGRAPHY STATE OCEANI C A.
PA   (SINO-) SINOGENOMAX CO LTD.
XX
PI   Xu X,  Yang F,  He J,  Pham L,  He M,  Ye Y,  Shen Y,  Kodira C;
XX
DR   WPI; 2001-355877/37.
DR   N-PSDB; AAH62710.
XX
PT   Primary nucleotide sequence of the shrimp white spot Bacilliform virus
PT   (WSBV), useful for producing viral polypeptides that can be used to
PT   screen for agents that are useful for treating WSBV infection.
XX
PS   Claim 1; Fig 3; 626pp; English.
XX
CC   The invention provides the primary nucleotide sequence of the WSBV genome
CC   (AAH62689), predicted transcript sequences (AAH62689-AAH62839) and
CC   encoded proteins (AAG84910-AAG85051) and oligonucleotide sequences
CC   (AAH62840-63160) suitable for use as primers or probes. The nucleic acid
CC   molecules and proteins of the invention are useful for diagnosis and
CC   monitoring viral infection, in screens for antiviral agents and for
CC   monitoring viral gene expression or activity during a treatment regimen.
CC   The nucleic acid molecules are also useful as antisense constructs to
CC   control viral gene expression in infected cells and tissues and to create
CC   transgenic viral resistant shrimp. (Updated on 06-AUG-2003 to correct OS
CC   field.)
XX
SQ   Sequence 1100 AA;

  Query Match             1.7%;  Score 11;  DB 4;  Length 1100;
  Best Local Similarity   100.0%;  Pred. No. 0.46;
  Matches   11;  Conservative    0;  Mismatches    0;  Indels    0;  Gaps    0;

Qy        380 SENEEEEEEEE 390
              ||||||||||
Db       1000 SENEEEEEEEE 1010


RESULT 10
AAG85039
ID   AAG85039 standard; protein; 1174 AA.
XX
AC   AAG85039;
XX
DT   06-AUG-2003  (revised)
DT   11-SEP-2001  (first entry)
XX
DE   Shrimp white spot Bacilliform virus (WSBV) protein 130.
XX
KW   Shrimp white spot Bacilliform virus; WSBV; diagnosis; viral infection;
KW   antiviral agent; gene expression; antisense construct;
KW   transgenic viral resistant shrimp.
XX
OS   White spot syndrome virus.
XX
PN   WO200138351-A2.
XX
PD   31-MAY-2001.
XX
PF   08-NOV-2000; 2000WO-US028888.
XX
PR   24-NOV-1999;   99CN-00124717.
XX
PA   (PENY-) PE CORP NY.
PA   (THIR-) THIRD INST OCEANOGRAPHY STATE OCEANI C A.
PA   (SINO-) SINOGENOMAX CO LTD.
XX
PI   Xu X,  Yang F,  He J,  Pham L,  He M,  Ye Y,  Shen Y,  Kodira C;
XX
DR   WPI; 2001-355877/37.
DR   N-PSDB; AAH62819.
XX
PT   Primary nucleotide sequence of the shrimp white spot Bacilliform virus
PT   (WSBV), useful for producing viral polypeptides that can be used to
PT   screen for agents that are useful for treating WSBV infection.
XX
PS   Claim 1; Fig 3; 626pp; English.
XX
CC   The invention provides the primary nucleotide sequence of the WSBV genome
CC   (AAH62689), predicted transcript sequences (AAH62689-AAH62839) and
CC   encoded proteins (AAG84910-AAG85051) and oligonucleotide sequences
CC   (AAH62840-63160) suitable for use as primers or probes. The nucleic acid
CC   molecules and proteins of the invention are useful for diagnosis and
CC   monitoring viral infection, in screens for antiviral agents and for
CC   monitoring viral gene expression or activity during a treatment regimen.
CC   The nucleic acid molecules are also useful as antisense constructs to
CC   control viral gene expression in infected cells and tissues and to create
CC   transgenic viral resistant shrimp. (Updated on 06-AUG-2003 to correct OS
CC   field.)
XX
SQ   Sequence 1174 AA;

  Query Match             1.7%;  Score 11;  DB 4;  Length 1174;
  Best Local Similarity   100.0%;  Pred. No. 0.48;
  Matches   11;  Conservative    0;  Mismatches    0;  Indels    0;  Gaps    0;

Qy        384 EEEEEEEDYED 394
              ||||||||||
Db         45 EEEEEEEDYED 55


RESULT 11
ADF69140
ID   ADF69140 standard; protein; 1483 AA.
XX
```

AC ADF69140;
XX
DT 12-FEB-2004 (first entry)
XX
DE Human MP53 protein sequence SEQ ID NO:110.
XX
KW p53 pathway modulating agent; MP53; p53 modulator; cytostatic;
KW gene therapy; cancer; human.
XX
OS Homo sapiens.
XX
PN WO2003083047-A2.
XX
PD 09-OCT-2003.
XX
PF 28-FEB-2003; 2003WO-US006025.
XX
PR 01-MAR-2002; 2002US-0361196P.
XX
PA (EXEL-) EXELIXIS INC.
XX
PI Belvin M, Francis-Lang H, Friedman L, Plowman GD, Heuer TS, Li D;
PI Funke RP;
XX
DR WPI; 2003-812540/76.
DR N-PSDB; ADF69196.
XX
PT Identifying a candidate p53 pathway modulating agent for treating e.g.,
PT cancer by contacting an assay system comprising a MP53 polypeptide or
PT nucleic acid with a test agent and detecting a test agent-biased
PT activity.
XX
PS Example; SEQ ID NO 110; 406pp; English.
XX
CC The present invention describes a method for identifying a candidate p53
CC pathway modulating agent, which comprises: (a) providing an assay system
CC comprising a MP53 (modulator of p53) polypeptide or nucleic acid or its
CC fragment or derivative; (b) contacting the assay system with a test agent
CC under conditions where the system provides a reference activity except in
CC the presence of the test agent; and (c) detecting a test agent-biased
CC activity, where a difference between the test agent-biased activity and
CC the reference activity identifies the test agent as a candidate p53
CC pathway modulating agent. Also described: (1) modulating the p53 pathway
CC of a cell; (2) modulating the p53 pathway in a mammalian cell; and (3)
CC diagnosing a disease in a patient. MP53 has cytostatic activity, and can
CC be used in gene therapy. The method is useful for identifying a candidate
CC p53 pathway modulating agent for preparing a composition for diagnosing
CC or treating e.g., cancer. The present sequence represents a human MP53
CC protein, which is used in the exemplification of the present invention.
XX
SQ Sequence 1483 AA;

Query Match 1.7%; Score 11; DB 7; Length 1483;
Best Local Similarity 100.0%; Pred. No. 0.59;
Matches 11; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

Qy 383 EEEEEEEEDYE 393
||||||||||||
Db 1266 EEEEEEEEDYE 1276

RESULT 12
ADQ20910
ID ADQ20910 standard; protein; 1483 AA.
XX
AC ADQ20910;
XX
DT 26-AUG-2004 (first entry)
XX
DE Human soft tissue sarcoma-upregulated protein - SEQ ID 3730.
XX
KW soft tissue sarcoma; cytostatic; gene therapy; vaccine; screening; human.
XX
OS Homo sapiens.
XX
PN WO2004048938-A2.
XX
PD 10-JUN-2004.
XX
PF 26-NOV-2003; 2003WO-US038193.
XX
PR 26-NOV-2002; 2002US-0429739P.
XX
PA (PROT-) PROTEIN DESIGN LABS INC.
XX
PI Aziz N, Ginsburg WM, Zlotnik A;
XX
DR WPI; 2004-441208/41.
XX
PT Early detection of soft tissue sarcoma comprises determining expression
PT of a gene in a first soft tissue sample and a normal soft tissue sample
PT and comparing the gene expression, also useful in treating soft tissue
PT sarcoma.
XX
PS Example 2; SEQ ID NO 3730; 210pp; English.
XX
CC The invention relates to a novel method for detecting soft tissue sarcoma
CC which comprises obtaining a first soft tissue sample from an individual
CC and a normal soft tissue sample from the same or different individual,
CC determining the expression of a gene in both samples and comparing the
CC expression of the gene in both soft tissue samples, where a higher level
CC of protein expression in the first soft tissue sample indicates the
CC presence of soft tissue sarcoma. The method of the invention has
CC cytostatic applications and may be useful for detecting soft tissue
CC sarcoma, possibly via gene therapy or vaccine production. The nucleic
CC acid sequences may be useful in diagnostic and screening applications.
CC The current sequence is that of a human soft tissue sarcoma-upregulated
CC protein of the invention. The current sequence is not shown within the
CC specification per se but was submitted in CD format by the inventor.
XX
SQ Sequence 1483 AA;

Query Match 1.7%; Score 11; DB 8; Length 1483;
Best Local Similarity 100.0%; Pred. No. 0.59;
Matches 11; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

Qy 383 EEEEEEEEDYE 393
||||||||||||
Db 1266 EEEEEEEEDYE 1276

RESULT 13
ADX05853
ID ADX05853 standard; protein; 1483 AA.
XX
AC ADX05853;
XX
DT 21-APR-2005 (first entry)
XX
DE Cyclin-dependent kinase modulation biomarker SEQ ID NO 418.
XX
KW cytostatic; cyclin-dependent kinase; cdk; biomarker.
XX
OS Homo sapiens.
XX
PN WO2005012875-A2.
XX
PD 10-FEB-2005.
XX
PF 29-JUL-2004; 2004WO-US024424.
XX
PR 29-JUL-2003; 2003US-0490890P.
XX
PA (BRIM) BRISTOL-MYERS SQUIBB CO.
XX
PI Li M, Rupnow BA, Webster KR, Jackson DG, Wong TW;

```
XX
DR   WPI; 2005-163068/17.
DR   N-PSDB; ADX05852.
XX
PT   Biomarkers useful for predicting or determining the response of a mammal
PT   to a cancer treatment comprising administration of a modulator of cyclin-
PT   dependent kinase activity.
XX
PS   Claim 5; SEQ ID NO 418; 141pp; English.
XX
CC   This invention describes a novel method of predicting or determining
CC   whether a mammal will respond or is responding to an anti-cancer agent
CC   that modulates cyclin-dependent kinase (cdk) activity. The method
CC   comprises measuring the level of one or more biomarkers selected from
CC   2774 biomarkers given in the specification (nucleotide sequence SEQ ID
CC   NO:1246 (Genbank EST W28729) is especially preferred). The method of the
CC   invention is utilized in a kit for determining or predicting whether
CC   patient would be susceptible or resistant to treatment by an agent
CC   modulating cdk activity. The invention also describes a method for
CC   utilizing individualized genetic profiles for treating diseases and
CC   disorders based on patient's response and molecular level, specialized
CC   microarrays comprising the biomarkers described, antibodies directed
CC   against the biomarkers and a cell culture model to identify biomarkers.
CC   The cdk modulator is preferably N-5-[[5-(1,1-Dimethylethyl)-2-
CC   oxazolyl]methyl]thio]-2-thiazolyl-4-piperidine carboxamide, 0.5-L-
CC   tartaric acid salt. Note: The sequence data for this patent did not form
CC   part of the printed specification, but was obtained in electronic format
CC   directly from WIPO at ftp.wipo.int/pub/published_pct_sequences. This
CC   sequence represents a biomarker used in the method of the invention.
XX
SQ   Sequence 1483 AA;

  Query Match             1.7%;  Score 11;  DB 9;  Length 1483;
  Best Local Similarity   100.0%;  Pred. No. 0.59;
  Matches   11;  Conservative    0;  Mismatches    0;  Indels    0;  Gaps    0;

Qy        383 EEEEEEEEDYE 393
              ||||||||||| 
Db       1266 EEEEEEEEDYE 1276


RESULT 14
ADX05855
ID   ADX05855 standard; protein; 1483 AA.
XX
AC   ADX05855;
XX
DT   21-APR-2005  (first entry)
XX
DE   Cyclin-dependent kinase modulation biomarker SEQ ID NO 420.
XX
KW   cytostatic; cyclin-dependent kinase; cdk; biomarker.
XX
OS   Homo sapiens.
XX
PN   WO2005012875-A2.
XX
PD   10-FEB-2005.
XX
PF   29-JUL-2004; 2004WO-US024424.
XX
PR   29-JUL-2003; 2003US-0490890P.
XX
PA   (BRIM ) BRISTOL-MYERS SQUIBB CO.
XX
PI   Li M,  Rupnow BA,  Webster KR,  Jackson DG,  Wong TW;
XX
DR   WPI; 2005-163068/17.
DR   N-PSDB; ADX05854.
XX
PT   Biomarkers useful for predicting or determining the response of a mammal
PT   to a cancer treatment comprising administration of a modulator of cyclin-
PT   dependent kinase activity.
XX
PS   Claim 5; SEQ ID NO 420; 141pp; English.
XX
CC   This invention describes a novel method of predicting or determining
CC   whether a mammal will respond or is responding to an anti-cancer agent
CC   that modulates cyclin-dependent kinase (cdk) activity. The method
CC   comprises measuring the level of one or more biomarkers selected from
CC   2774 biomarkers given in the specification (nucleotide sequence SEQ ID
CC   NO:1246 (Genbank EST W28729) is especially preferred). The method of the
CC   invention is utilized in a kit for determining or predicting whether
CC   patient would be susceptible or resistant to treatment by an agent
CC   modulating cdk activity. The invention also describes a method for
CC   utilizing individualized genetic profiles for treating diseases and
CC   disorders based on patient's response and molecular level, specialized
CC   microarrays comprising the biomarkers described, antibodies directed
CC   against the biomarkers and a cell culture model to identify biomarkers.
CC   The cdk modulator is preferably N-5-[[5-(1,1-Dimethylethyl)-2-
CC   oxazolyl]methyl]thio]-2-thiazolyl-4-piperidine carboxamide, 0.5-L-
CC   tartaric acid salt. Note: The sequence data for this patent did not form
CC   part of the printed specification, but was obtained in electronic format
CC   directly from WIPO at ftp.wipo.int/pub/published_pct_sequences. This
CC   sequence represents a biomarker used in the method of the invention.
XX
SQ   Sequence 1483 AA;

  Query Match             1.7%;  Score 11;  DB 9;  Length 1483;
  Best Local Similarity   100.0%;  Pred. No. 0.59;
  Matches   11;  Conservative    0;  Mismatches    0;  Indels    0;  Gaps    0;

Qy        383 EEEEEEEEDYE 393
              ||||||||||| 
Db       1266 EEEEEEEEDYE 1276


RESULT 15
AAW81172
ID   AAW81172 standard; protein; 1527 AA.
XX
AC   AAW81172;
XX
DT   05-MAY-1999  (first entry)
XX
DE   Human BAZ1-beta protein #1.
XX
KW   Transcriptional regulator; transcription; BAZ1-alpha; bromodomain; BAZ;
KW   atypical zinc finger; testis; human; tumour; BAZ1-beta; BAZ2-alpha; drug;
KW   BAZ2-beta; treatment; cancer; proliferative disorder; screening.
XX
OS   Homo sapiens.
XX
PN   WO9847920-A1.
XX
PD   29-OCT-1998.
XX
PF   17-APR-1998;   98WO-JP001783.
XX
PR   18-APR-1997;   97JP-00116570.
PR   24-OCT-1997;   97JP-00310027.
XX
PA   (CHUS ) CHUGAI RES INST MOLECULAR MEDICINE INC.
XX
PI   Jones MH;
XX
DR   WPI; 1998-583603/49.
DR   N-PSDB; AAV68408.
XX
PT   Transcriptional regulator gene family containing bromodomain - may be
PT   expressed in testis tissue and is useful for treatment of cancer and
PT   other proliferative disorders.
XX
PS   Claim 1; Page 125-137; 187pp; Japanese.
XX
CC   This sequence represents the human BAZ1-beta protein, a member of a
```

CC family of transcriptional regulator genes containing a bromodomain (BAZ, CC Bromodomain with Atypical Zinc finger) which are expressed specifically CC in testis tissue and also in certain tumour lines. Transgenic cells may CC be used for the preparation of the BAZ1-alpha, BAZ1-beta, BAZ2-alpha and CC BAZ2-beta proteins. These proteins can be used in the treatment of cancer CC and other proliferative disorders, and in screening of compounds for CC their binding ability to the expression products (e.g. for use as drugs CC by modulation of transcriptional regulation)
XX
SQ Sequence 1527 AA;

Query Match 1.7%; Score 11; DB 2; Length 1527;
Best Local Similarity 100.0%; Pred. No. 0.61;
Matches 11; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

```
Qy        383 EEEEEEEEDYE 393
              ||||||||||| 
Db       1266 EEEEEEEEDYE 1276
```

Search completed: February 28, 2006, 09:05:50
Job time : 188 secs

OM protein - protein search, using sw model

Run on: February 28, 2006, 09:10:01 ; Search time 47 Seconds
(without alignments)
1124.037 Million cell updates/sec

Title: US-10-717-665A-44
Perfect score: 639
Sequence: 1 MPQPSVSGMDPPFGDAFRSH.....TAEGNPTGGLVGLRIPTSKV 639

Scoring table: OLIGO
Gapop 60.0 , Gapext 60.0

Searched: 572060 seqs, 82675679 residues

Word size : 0

Total number of hits satisfying chosen parameters: 572060

Minimum DB seq length: 0
Maximum DB seq length: 2000000000

Post-processing: Listing first 45 summaries

Database : Issued_Patents_AA:*
1: /cgn2_6/ptodata/1/iaa/5_COMB.pep:*
2: /cgn2_6/ptodata/1/iaa/6_COMB.pep:*
3: /cgn2_6/ptodata/1/iaa/H_COMB.pep:*
4: /cgn2_6/ptodata/1/iaa/PCTUS_COMB.pep:*
5: /cgn2_6/ptodata/1/iaa/RE_COMB.pep:*
6: /cgn2_6/ptodata/1/iaa/backfiles1.pep:*

Pred. No. is the number of results predicted by chance to have a score greater than or equal to the score of the result being printed, and is derived by analysis of the total score distribution.

SUMMARIES

Result No.	Score	% Query Match	Length	DB	ID	Description
1	639	100.0	639	2	US-10-164-595-44	Sequence 44, Appl
2	83	13.0	91	2	US-10-164-595-75	Sequence 75, Appl
3	11	1.7	830	2	US-09-562-737-35	Sequence 35, Appl
4	11	1.7	1525	2	US-09-418-710-69	Sequence 69, Appl
5	11	1.7	1525	2	US-09-839-479-68	Sequence 68, Appl
6	11	1.7	1527	2	US-09-418-710-27	Sequence 27, Appl
7	11	1.7	1527	2	US-09-839-479-27	Sequence 27, Appl
8	11	1.7	1531	2	US-09-418-710-29	Sequence 29, Appl
9	11	1.7	1531	2	US-09-839-479-29	Sequence 29, Appl
10	11	1.7	1540	2	US-09-949-016-7037	Sequence 7037, Ap
11	10	1.6	114	2	US-09-248-796A-23592	Sequence 23592, A
12	10	1.6	141	2	US-09-270-767-41799	Sequence 41799, A
13	10	1.6	141	2	US-09-270-767-57041	Sequence 57041, A
14	10	1.6	236	2	US-10-202-480-4	Sequence 4, Appli
15	10	1.6	255	2	US-09-991-181-287	Sequence 287, App
16	10	1.6	255	2	US-09-990-444-287	Sequence 287, App
17	10	1.6	255	2	US-09-997-333-287	Sequence 287, App
18	10	1.6	255	2	US-09-992-598-287	Sequence 287, App
19	10	1.6	273	2	US-10-202-480-2	Sequence 2, Appli
20	10	1.6	289	2	US-09-792-024-108	Sequence 108, App
21	10	1.6	639	2	US-09-248-796A-17567	Sequence 17567, A
22	9	1.4	16	2	US-08-979-608A-25	Sequence 25, Appl
23	9	1.4	16	2	US-09-517-849-25	Sequence 25, Appl
24	9	1.4	16	2	US-09-616-289-25	Sequence 25, Appl
25	9	1.4	16	2	US-09-976-740-25	Sequence 25, Appl
26	9	1.4	28	2	US-08-979-608A-26	Sequence 26, Appl
27	9	1.4	28	2	US-09-517-849-26	Sequence 26, Appl
28	9	1.4	28	2	US-09-616-289-26	Sequence 26, Appl
29	9	1.4	28	2	US-09-976-740-26	Sequence 26, Appl
30	9	1.4	62	2	US-09-248-796A-22544	Sequence 22544, A
31	9	1.4	67	2	US-09-018-635-33	Sequence 33, Appl
32	9	1.4	67	2	US-09-912-962-33	Sequence 33, Appl
33	9	1.4	68	2	US-09-513-999C-6304	Sequence 6304, Ap
34	9	1.4	134	2	US-09-513-999C-6305	Sequence 6305, Ap
35	9	1.4	137	2	US-09-621-976-5741	Sequence 5741, Ap
36	9	1.4	141	2	US-09-621-976-6363	Sequence 6363, Ap
37	9	1.4	156	1	US-08-552-142A-9	Sequence 9, Appli
38	9	1.4	156	1	US-08-910-973-9	Sequence 9, Appli
39	9	1.4	156	2	US-09-499-227-9	Sequence 9, Appli
40	9	1.4	156	4	PCT-US95-05741-9	Sequence 9, Appli
41	9	1.4	175	2	US-09-248-796A-14111	Sequence 14111, A
42	9	1.4	175	2	US-09-248-796A-24049	Sequence 24049, A
43	9	1.4	200	2	US-09-702-705-324	Sequence 324, App
44	9	1.4	200	2	US-09-702-705-789	Sequence 789, App
45	9	1.4	200	2	US-09-736-457-324	Sequence 324, App

ALIGNMENTS

RESULT 1
US-10-164-595-44
; Sequence 44, Application US/10164595
; Patent No. 6657054
; GENERAL INFORMATION:
; APPLICANT: OriGene Technologies, Inc
; TITLE OF INVENTION: Regulated Angiogenesis Genes and Polypeptides
; FILE REFERENCE: 1U 103 R1
; CURRENT APPLICATION NUMBER: US/10/164,595
; CURRENT FILING DATE: 2002-06-10
; NUMBER OF SEQ ID NOS: 80
; SOFTWARE: PatentIn version 3.1
; SEQ ID NO 44
; LENGTH: 639
; TYPE: PRT
; ORGANISM: Homo sapiens
US-10-164-595-44

Query Match 100.0%; Score 639; DB 2; Length 639;
Best Local Similarity 100.0%; Pred. No. 0;
Matches 639; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

```
Qy          1 MPQPSVSGMDPPFGDAFRSHTFSEQTLMSTDLLANSSDPDFMYELDREMNYQQNPRDNFL 60
              ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db          1 MPQPSVSGMDPPFGDAFRSHTFSEQTLMSTDLLANSSDPDFMYELDREMNYQQNPRDNFL 60

Qy         61 SLEDCKDIENLESFTDVLDNEGALTSNWEQWDTYCEDLTKYTKLTSCDIWGTKEVDYLGL 120
              ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db         61 SLEDCKDIENLESFTDVLDNEGALTSNWEQWDTYCEDLTKYTKLTSCDIWGTKEVDYLGL 120

Qy        121 DDFSSPYQDEEVISKTPTLAQLNSEDSQSVSDSLYYPDSLFSVKQNPLPSSFPGKKITSR 180
              ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db        121 DDFSSPYQDEEVISKTPTLAQLNSEDSQSVSDSLYYPDSLFSVKQNPLPSSFPGKKITSR 180

Qy        181 AAAPVCSSKTLQAEVPLSDCVQKASKPPSSTQIMVKTNMYHNEKVNFHVECKDYVKKAKV 240
              ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db        181 AAAPVCSSKTLQAEVPLSDCVQKASKPPSSTQIMVKTNMYHNEKVNFHVECKDYVKKAKV 240

Qy        241 KINPVQQSRPLLSQIHTDAAKENTCYCGAVAKRQEKKGMEPLQGHATPALPFKETQELLL 300
              ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db        241 KINPVQQSRPLLSQIHTDAAKENTCYCGAVAKRQEKKGMEPLQGHATPALPFKETQELLL 300

Qy        301 SPLPQEGPGSLAAGESSSLSASTSVSDSSQKKEEHNYSLFVSDNLGEQPTKCSPEEDEED 360
              ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db        301 SPLPQEGPGSLAAGESSSLSASTSVSDSSQKKEEHNYSLFVSDNLGEQPTKCSPEEDEED 360

Qy        361 EEDVDDEDHDEGFGSEHELSENEEEEEEEEDYEDDKDDDISDTFSEPGYENDSVEDLKEV 420
              ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db        361 EEDVDDEDHDEGFGSEHELSENEEEEEEEEDYEDDKDDDISDTFSEPGYENDSVEDLKEV 420
```

```
Qy        421 TSISSRKRGKRRYFWEYSEQLTPSQQERMLRPSEWNRDTLPSNMYQKNGLHHGKYAVKKS 480
              ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db        421 TSISSRKRGKRRYFWEYSEQLTPSQQERMLRPSEWNRDTLPSNMYQKNGLHHGKYAVKKS 480

Qy        481 RRTDVEDLTPNPKKLLQIGNELRKLNKVISDLTPVSELPLTARPRSRKEKNKLAFRACRL 540
              ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db        481 RRTDVEDLTPNPKKLLQIGNELRKLNKVISDLTPVSELPLTARPRSRKEKNKLAFRACRL 540

Qy        541 KKKAQYEANKVKLWGLNTEYDNLLFVINSIKQEIVNRVQNPRDERGPNMGQKLEILIKDT 600
              ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db        541 KKKAQYEANKVKLWGLNTEYDNLLFVINSIKQEIVNRVQNPRDERGPNMGQKLEILIKDT 600

Qy        601 LGLPVAGQTSEFVNQVLEKTAEGNPTGGLVGLRIPTSKV 639
              |||||||||||||||||||||||||||||||||||||||
Db        601 LGLPVAGQTSEFVNQVLEKTAEGNPTGGLVGLRIPTSKV 639


RESULT 2
US-10-164-595-75
; Sequence 75, Application US/10164595
; Patent No. 6657054
; GENERAL INFORMATION:
;  APPLICANT: OriGene Technologies, Inc
;  TITLE OF INVENTION: Regulated Angiogenesis Genes and Polypeptides
;  FILE REFERENCE: 1U 103 R1
;  CURRENT APPLICATION NUMBER: US/10/164,595
;  CURRENT FILING DATE:  2002-06-10
;  NUMBER OF SEQ ID NOS: 80
;  SOFTWARE: PatentIn version 3.1
; SEQ ID NO 75
;   LENGTH: 91
;   TYPE: PRT
;   ORGANISM: Homo sapiens
US-10-164-595-75

  Query Match             13.0%;  Score 83;  DB 2;  Length 91;
  Best Local Similarity   100.0%;  Pred. No. 1.4e-70;
  Matches   83;  Conservative    0;  Mismatches    0;  Indels    0;  Gaps    0;

Qy        449 MLRPSEWNRDTLPSNMYQKNGLHHGKYAVKKSRRTDVEDLTPNPKKLLQIGNELRKLNKV 508
              ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db          1 MLRPSEWNRDTLPSNMYQKNGLHHGKYAVKKSRRTDVEDLTPNPKKLLQIGNELRKLNKV 60

Qy        509 ISDLTPVSELPLTARPRSRKEKN 531
              |||||||||||||||||||||||
Db         61 ISDLTPVSELPLTARPRSRKEKN 83


RESULT 3
US-09-562-737-35
; Sequence 35, Application US/09562737
; Patent No. 6428967
; GENERAL INFORMATION:
;  APPLICANT:  Herz, Joachim
;  APPLICANT:  Gotthardt, Michael
;  TITLE OF INVENTION:  LDL Receptor Signaling Pathways
;  FILE REFERENCE:  UTSW0708
;  CURRENT APPLICATION NUMBER: US/09/562,737
;  CURRENT FILING DATE:  2000-05-01
;  NUMBER OF SEQ ID NOS:  132
;  SOFTWARE:  PatentIn Ver. 2.1
; SEQ ID NO 35
;   LENGTH: 830
;   TYPE: PRT
;   ORGANISM: Artificial Sequence
;   FEATURE:
;   OTHER INFORMATION: Description of Artificial Sequence: Synthetic
;   OTHER INFORMATION: Sequence
US-09-562-737-35

  Query Match             1.7%;  Score 11;  DB 2;  Length 830;
  Best Local Similarity   100.0%;  Pred. No. 0.15;
  Matches   11;  Conservative    0;  Mismatches    0;  Indels    0;  Gaps    0;

Qy        381 ENEEEEEEEED 391
              ||||||||||||
Db         94 ENEEEEEEEED 104


RESULT 4
US-09-418-710-69
; Sequence 69, Application US/09418710
; Patent No. 6596482
; GENERAL INFORMATION:
;  APPLICANT:  Jones, Michael H.
;  TITLE OF INVENTION:  TRANSCRIPTIONAL REGULATOR
;  FILE REFERENCE:  06501-042001
;  CURRENT APPLICATION NUMBER: US/09/418,710
;  CURRENT FILING DATE:  1999-10-15
;  PRIOR APPLICATION NUMBER:  PCT/JP98/01783
;  PRIOR FILING DATE:  1998-04-17
;  PRIOR APPLICATION NUMBER:  JP 9/310027
;  PRIOR FILING DATE:  1997-10-24
;  PRIOR APPLICATION NUMBER:  JP 9/116570
;  PRIOR FILING DATE:  1997-04-18
;  NUMBER OF SEQ ID NOS:  73
;  SOFTWARE:  FastSEQ for Windows Version 4.0
; SEQ ID NO 69
;   LENGTH: 1525
;   TYPE: PRT
;   ORGANISM: Homo sapiens
US-09-418-710-69

  Query Match             1.7%;  Score 11;  DB 2;  Length 1525;
  Best Local Similarity   100.0%;  Pred. No. 0.25;
  Matches   11;  Conservative    0;  Mismatches    0;  Indels    0;  Gaps    0;

Qy        383 EEEEEEEEDYE 393
              ||||||||||||
Db       1264 EEEEEEEEDYE 1274


RESULT 5
US-09-839-479-68
; Sequence 68, Application US/09839479
; Patent No. 6727222
; GENERAL INFORMATION:
;  APPLICANT: Jones, Michael H.
;  TITLE OF INVENTION: TRANSCRIPTIONAL REGULATOR
;  FILE REFERENCE: 06501-042002
;  CURRENT APPLICATION NUMBER: US/09/839,479
;  CURRENT FILING DATE:  2001-04-20
;  PRIOR APPLICATION NUMBER: US 09/418,710
;  PRIOR FILING DATE: 1999-10-15
;  PRIOR APPLICATION NUMBER: PCT/JP98/01783
;  PRIOR FILING DATE: 1998-04-17
;  PRIOR APPLICATION NUMBER: JP 9/310027
;  PRIOR FILING DATE: 1997-10-24
;  PRIOR APPLICATION NUMBER: JP 9/116570
;  PRIOR FILING DATE: 1997-04-18
;  NUMBER OF SEQ ID NOS: 72
;  SOFTWARE: FastSEQ for Windows Version 4.0
; SEQ ID NO 68
;   LENGTH: 1525
;   TYPE: PRT
;   ORGANISM: Homo sapiens
US-09-839-479-68

  Query Match             1.7%;  Score 11;  DB 2;  Length 1525;
  Best Local Similarity   100.0%;  Pred. No. 0.25;
  Matches   11;  Conservative    0;  Mismatches    0;  Indels    0;  Gaps    0;

Qy        383 EEEEEEEEDYE 393
              ||||||||||||
Db       1264 EEEEEEEEDYE 1274
```

```
RESULT 6
US-09-418-710-27
; Sequence 27, Application US/09418710
; Patent No. 6596482
; GENERAL INFORMATION:
;  APPLICANT:  Jones, Michael H.
;  TITLE OF INVENTION:  TRANSCRIPTIONAL REGULATOR
;  FILE REFERENCE:  06501-042001
;  CURRENT APPLICATION NUMBER: US/09/418,710
;  CURRENT FILING DATE:  1999-10-15
;  PRIOR APPLICATION NUMBER:  PCT/JP98/01783
;  PRIOR FILING DATE:  1998-04-17
;  PRIOR APPLICATION NUMBER:  JP 9/310027
;  PRIOR FILING DATE:  1997-10-24
;  PRIOR APPLICATION NUMBER:  JP 9/116570
;  PRIOR FILING DATE:  1997-04-18
;  NUMBER OF SEQ ID NOS:  73
;  SOFTWARE:  FastSEQ for Windows Version 4.0
; SEQ ID NO 27
;   LENGTH: 1527
;   TYPE: PRT
;   ORGANISM: Homo sapiens
US-09-418-710-27

  Query Match             1.7%;  Score 11;  DB 2;  Length 1527;
  Best Local Similarity   100.0%;  Pred. No. 0.25;
  Matches   11;  Conservative    0;  Mismatches    0;  Indels    0;  Gaps    0;

Qy        383 EEEEEEEEDYE 393
              ||||||||||| 
Db       1266 EEEEEEEEDYE 1276


RESULT 7
US-09-839-479-27
; Sequence 27, Application US/09839479
; Patent No. 6727222
; GENERAL INFORMATION:
;  APPLICANT: Jones, Michael H.
;  TITLE OF INVENTION: TRANSCRIPTIONAL REGULATOR
;  FILE REFERENCE: 06501-042002
;  CURRENT APPLICATION NUMBER: US/09/839,479
;  CURRENT FILING DATE:  2001-04-20
;  PRIOR APPLICATION NUMBER: US 09/418,710
;  PRIOR FILING DATE: 1999-10-15
;  PRIOR APPLICATION NUMBER: PCT/JP98/01783
;  PRIOR FILING DATE: 1998-04-17
;  PRIOR APPLICATION NUMBER: JP 9/310027
;  PRIOR FILING DATE: 1997-10-24
;  PRIOR APPLICATION NUMBER: JP 9/116570
;  PRIOR FILING DATE: 1997-04-18
;  NUMBER OF SEQ ID NOS: 72
;  SOFTWARE: FastSEQ for Windows Version 4.0
; SEQ ID NO 27
;   LENGTH: 1527
;   TYPE: PRT
;   ORGANISM: Homo sapiens
US-09-839-479-27

  Query Match             1.7%;  Score 11;  DB 2;  Length 1527;
  Best Local Similarity   100.0%;  Pred. No. 0.25;
  Matches   11;  Conservative    0;  Mismatches    0;  Indels    0;  Gaps    0;

Qy        383 EEEEEEEEDYE 393
              ||||||||||| 
Db       1266 EEEEEEEEDYE 1276


RESULT 8
US-09-418-710-29
; Sequence 29, Application US/09418710
; Patent No. 6596482
; GENERAL INFORMATION:
;  APPLICANT:  Jones, Michael H.
;  TITLE OF INVENTION:  TRANSCRIPTIONAL REGULATOR
;  FILE REFERENCE:  06501-042001
;  CURRENT APPLICATION NUMBER: US/09/418,710
;  CURRENT FILING DATE:  1999-10-15
;  PRIOR APPLICATION NUMBER:  PCT/JP98/01783
;  PRIOR FILING DATE:  1998-04-17
;  PRIOR APPLICATION NUMBER:  JP 9/310027
;  PRIOR FILING DATE:  1997-10-24
;  PRIOR APPLICATION NUMBER:  JP 9/116570
;  PRIOR FILING DATE:  1997-04-18
;  NUMBER OF SEQ ID NOS:  73
;  SOFTWARE:  FastSEQ for Windows Version 4.0
; SEQ ID NO 29
;   LENGTH: 1531
;   TYPE: PRT
;   ORGANISM: Homo sapiens
US-09-418-710-29

  Query Match             1.7%;  Score 11;  DB 2;  Length 1531;
  Best Local Similarity   100.0%;  Pred. No. 0.25;
  Matches   11;  Conservative    0;  Mismatches    0;  Indels    0;  Gaps    0;

Qy        383 EEEEEEEEDYE 393
              ||||||||||| 
Db       1270 EEEEEEEEDYE 1280


RESULT 9
US-09-839-479-29
; Sequence 29, Application US/09839479
; Patent No. 6727222
; GENERAL INFORMATION:
;  APPLICANT: Jones, Michael H.
;  TITLE OF INVENTION: TRANSCRIPTIONAL REGULATOR
;  FILE REFERENCE: 06501-042002
;  CURRENT APPLICATION NUMBER: US/09/839,479
;  CURRENT FILING DATE:  2001-04-20
;  PRIOR APPLICATION NUMBER: US 09/418,710
;  PRIOR FILING DATE: 1999-10-15
;  PRIOR APPLICATION NUMBER: PCT/JP98/01783
;  PRIOR FILING DATE: 1998-04-17
;  PRIOR APPLICATION NUMBER: JP 9/310027
;  PRIOR FILING DATE: 1997-10-24
;  PRIOR APPLICATION NUMBER: JP 9/116570
;  PRIOR FILING DATE: 1997-04-18
;  NUMBER OF SEQ ID NOS: 72
;  SOFTWARE: FastSEQ for Windows Version 4.0
; SEQ ID NO 29
;   LENGTH: 1531
;   TYPE: PRT
;   ORGANISM: Homo sapiens
US-09-839-479-29

  Query Match             1.7%;  Score 11;  DB 2;  Length 1531;
  Best Local Similarity   100.0%;  Pred. No. 0.25;
  Matches   11;  Conservative    0;  Mismatches    0;  Indels    0;  Gaps    0;

Qy        383 EEEEEEEEDYE 393
              ||||||||||| 
Db       1270 EEEEEEEEDYE 1280


RESULT 10
US-09-949-016-7037
; Sequence 7037, Application US/09949016
; Patent No. 6812339
; GENERAL INFORMATION:
;  APPLICANT: VENTER, J. Craig et al.
;  TITLE OF INVENTION: POLYMORPHISMS IN KNOWN GENES ASSOCIATED
;  TITLE OF INVENTION:  WITH HUMAN DISEASE, METHODS OF DETECTION AND USES THEREOF
;  FILE REFERENCE: CL001307
```

```
;  CURRENT APPLICATION NUMBER: US/09/949,016
;  CURRENT FILING DATE:  2000-04-14
;  PRIOR APPLICATION NUMBER: 60/241,755
;  PRIOR FILING DATE: 2000-10-20
;  PRIOR APPLICATION NUMBER: 60/237,768
;  PRIOR FILING DATE: 2000-10-03
;  PRIOR APPLICATION NUMBER: 60/231,498
;  PRIOR FILING DATE: 2000-09-08
;  NUMBER OF SEQ ID NOS: 207012
;  SOFTWARE: FastSEQ for Windows Version 4.0
; SEQ ID NO 7037
;   LENGTH: 1540
;   TYPE: PRT
;   ORGANISM: Human
US-09-949-016-7037

  Query Match             1.7%;  Score 11;  DB 2;  Length 1540;
  Best Local Similarity   100.0%;  Pred. No. 0.25;
  Matches   11;  Conservative    0;  Mismatches    0;  Indels    0;  Gaps    0;

Qy        383 EEEEEEEEDYE 393
              ||||||||||
Db       1323 EEEEEEEEDYE 1333


RESULT 11
US-09-248-796A-23592
; Sequence 23592, Application US/09248796A
; Patent No. 6747137
; GENERAL INFORMATION:
;  APPLICANT: Keith Weinstock et al
;  TITLE OF INVENTION: NUCLEIC ACID AND AMINO ACID SEQUENCES RELATING TO CANDIDA ALBICAN
;  TITLE OF INVENTION:  FOR DIAGNOSTICS AND THERAPEUTICS
;  FILE REFERENCE: 107196.132
;  CURRENT APPLICATION NUMBER: US/09/248,796A
;  CURRENT FILING DATE:  1999-02-12
;  PRIOR APPLICATION NUMBER: US 60/074,725
;  PRIOR FILING DATE: 1998-02-13
;  PRIOR APPLICATION NUMBER: US 60/096,409
;  PRIOR FILING DATE: 1998-08-13
;  NUMBER OF SEQ ID NOS: 28208
; SEQ ID NO 23592
;   LENGTH: 114
;   TYPE: PRT
;   ORGANISM: Candida albicans
US-09-248-796A-23592

  Query Match             1.6%;  Score 10;  DB 2;  Length 114;
  Best Local Similarity   100.0%;  Pred. No. 0.22;
  Matches   10;  Conservative    0;  Mismatches    0;  Indels    0;  Gaps    0;

Qy        382 NEEEEEEEED 391
              ||||||||||
Db         95 NEEEEEEEED 104


RESULT 12
US-09-270-767-41799
; Sequence 41799, Application US/09270767
; Patent No. 6703491
; GENERAL INFORMATION:
;  APPLICANT: Homburger et al.
;  TITLE OF INVENTION: Nucleic acids and proteins of Drosophila melanogaster
;  FILE REFERENCE: File Reference: 7326-094
;  CURRENT APPLICATION NUMBER: US/09/270,767
;  CURRENT FILING DATE: 1999-03-17
;  NUMBER OF SEQ ID NOS: 62517
;  SOFTWARE: PatentIn Ver. 2.0
; SEQ ID NO 41799
;   LENGTH: 141
;   TYPE: PRT
;   ORGANISM: Drosophila melanogaster
US-09-270-767-41799

  Query Match             1.6%;  Score 10;  DB 2;  Length 141;
  Best Local Similarity   100.0%;  Pred. No. 0.26;
  Matches   10;  Conservative    0;  Mismatches    0;  Indels    0;  Gaps    0;

Qy        536 RACRLKKKAQ 545
              ||||||||||
Db         39 RACRLKKKAQ 48


RESULT 13
US-09-270-767-57041
; Sequence 57041, Application US/09270767
; Patent No. 6703491
; GENERAL INFORMATION:
;  APPLICANT: Homburger et al.
;  TITLE OF INVENTION: Nucleic acids and proteins of Drosophila melanogaster
;  FILE REFERENCE: File Reference: 7326-094
;  CURRENT APPLICATION NUMBER: US/09/270,767
;  CURRENT FILING DATE: 1999-03-17
;  NUMBER OF SEQ ID NOS: 62517
;  SOFTWARE: PatentIn Ver. 2.0
; SEQ ID NO 57041
;   LENGTH: 141
;   TYPE: PRT
;   ORGANISM: Drosophila melanogaster
US-09-270-767-57041

  Query Match             1.6%;  Score 10;  DB 2;  Length 141;
  Best Local Similarity   100.0%;  Pred. No. 0.26;
  Matches   10;  Conservative    0;  Mismatches    0;  Indels    0;  Gaps    0;

Qy        536 RACRLKKKAQ 545
              ||||||||||
Db         39 RACRLKKKAQ 48


RESULT 14
US-10-202-480-4
; Sequence 4, Application US/10202480
; Patent No. 6929923
; GENERAL INFORMATION:
;  APPLICANT: Vallone, Marcy K.
;  APPLICANT:  Wong, Brian R.
;  APPLICANT:  Masuda, Esteban
;  APPLICANT:  Powell, Mark
;  TITLE OF INVENTION: Modulators of Leukocyte Activation, BIC Compositions and Methods c
;  FILE REFERENCE: A-71314/RMS/DHR
;  CURRENT APPLICATION NUMBER: US/10/202,480
;  CURRENT FILING DATE:  2002-11-14
;  NUMBER OF SEQ ID NOS: 4
;  SOFTWARE: PatentIn version 3.1
; SEQ ID NO 4
;   LENGTH: 236
;   TYPE: PRT
;   ORGANISM: Homo sapiens
US-10-202-480-4

  Query Match             1.6%;  Score 10;  DB 2;  Length 236;
  Best Local Similarity   100.0%;  Pred. No. 0.41;
  Matches   10;  Conservative    0;  Mismatches    0;  Indels    0;  Gaps    0;

Qy        381 ENEEEEEEEE 390
              ||||||||||
Db         46 ENEEEEEEEE 55


RESULT 15
US-09-991-181-287
; Sequence 287, Application US/09991181
; Patent No. 6913919
; GENERAL INFORMATION:
;  APPLICANT: Ashkenazi,Avi J.
```

; APPLICANT: Baker,Kevin P.
; APPLICANT: Botstein,David
; APPLICANT: Desnoyers,Luc
; APPLICANT: Eaton,Dan L.
; APPLICANT: Ferrara,Napoleone
; APPLICANT: Fong,Sherman
; APPLICANT: Gerber,Hanspeter
; APPLICANT: Gerritsen,Mary E.
; APPLICANT: Goddard,Audrey
; APPLICANT: Godowski,Paul J.
; APPLICANT: Grimaldi,J.Christopher
; APPLICANT: Gurney,Austin L.
; APPLICANT: Kljavin,Ivar J.
; APPLICANT: Napier,Mary A.
; APPLICANT: Pan,James
; APPLICANT: Paoni,Nicholas F.
; APPLICANT: Roy, Margaret Ann
; APPLICANT: Stewart, Timothy A.
; APPLICANT: Tumas, Daniel
; APPLICANT: Watanabe, Colin K.
; APPLICANT: Williams, P. Mickey
; APPLICANT: Wood, William I.
; APPLICANT: Zhang, Zemin
; TITLE OF INVENTION: Secreted and Transmembrane Polypeptides and Nucleic
; TITLE OF INVENTION: Acids Encoding the Same
; FILE REFERENCE: P2730P1C53
; CURRENT APPLICATION NUMBER: US/09/991,181
; CURRENT FILING DATE: 2001-11-16
; PRIOR APPLICATION NUMBER: 60/049787
; PRIOR FILING DATE: 1997-06-16
; PRIOR APPLICATION NUMBER: 60/062250
; PRIOR FILING DATE: 1997-10-17
; PRIOR APPLICATION NUMBER: 60/065186
; PRIOR FILING DATE: 1997-11-12
; PRIOR APPLICATION NUMBER: 60/065311
; PRIOR FILING DATE: 1997-11-13
; PRIOR APPLICATION NUMBER: 60/066770
; PRIOR FILING DATE: 1997-11-24
; PRIOR APPLICATION NUMBER: 60/075945
; PRIOR FILING DATE: 1998-02-25
; PRIOR APPLICATION NUMBER: 60/078910
; PRIOR FILING DATE: 1998-03-20
; PRIOR APPLICATION NUMBER: 60/083322
; PRIOR FILING DATE: 1998-04-28
; PRIOR APPLICATION NUMBER: 60/084600
; PRIOR FILING DATE: 1998-05-07
; PRIOR APPLICATION NUMBER: 60/087106
; PRIOR FILING DATE: 1998-05-28
; PRIOR APPLICATION NUMBER: 60/087607
; PRIOR FILING DATE: 1998-06-02
; PRIOR APPLICATION NUMBER: 60/087609
; PRIOR FILING DATE: 1998-06-02
; PRIOR APPLICATION NUMBER: 60/087759
; PRIOR FILING DATE: 1998-06-02
; PRIOR APPLICATION NUMBER: 60/087827
; PRIOR FILING DATE: 1998-06-03
; PRIOR APPLICATION NUMBER: 60/088021
; PRIOR FILING DATE: 1998-06-04
; PRIOR APPLICATION NUMBER: 60/088025
; PRIOR FILING DATE: 1998-06-04
; PRIOR APPLICATION NUMBER: 60/088026
; PRIOR FILING DATE: 1998-06-04
; PRIOR APPLICATION NUMBER: 60/088028
; PRIOR FILING DATE: 1998-06-04
; PRIOR APPLICATION NUMBER: 60/088029
; PRIOR FILING DATE: 1998-06-04
; PRIOR APPLICATION NUMBER: 60/088030
; PRIOR FILING DATE: 1998-06-04
; PRIOR APPLICATION NUMBER: 60/088033
; PRIOR FILING DATE: 1998-06-04
; PRIOR APPLICATION NUMBER: 60/088326
; PRIOR FILING DATE: 1998-06-04
; PRIOR APPLICATION NUMBER: 60/088167
; PRIOR FILING DATE: 1998-06-05
; PRIOR APPLICATION NUMBER: 60/088202
; PRIOR FILING DATE: 1998-06-05
; PRIOR APPLICATION NUMBER: 60/088212
; PRIOR FILING DATE: 1998-06-05
; PRIOR APPLICATION NUMBER: 60/088217
; PRIOR FILING DATE: 1998-06-05
; PRIOR APPLICATION NUMBER: 60/088655
; PRIOR FILING DATE: 1998-06-09
; PRIOR APPLICATION NUMBER: 60/088734
; PRIOR FILING DATE: 1998-06-10
; PRIOR APPLICATION NUMBER: 60/088738
; PRIOR FILING DATE: 1998-06-10
; PRIOR APPLICATION NUMBER: 60/088742
; PRIOR FILING DATE: 1998-06-10
; PRIOR APPLICATION NUMBER: 60/088810
; PRIOR FILING DATE: 1998-06-10
; PRIOR APPLICATION NUMBER: 60/088824
; PRIOR FILING DATE: 1998-06-10
; PRIOR APPLICATION NUMBER: 60/088826
; PRIOR FILING DATE: 1998-06-10
; PRIOR APPLICATION NUMBER: 60/088858
; PRIOR FILING DATE: 1998-06-11
; PRIOR APPLICATION NUMBER: 60/088861
; PRIOR FILING DATE: 1998-06-11
; PRIOR APPLICATION NUMBER: 60/088876
; PRIOR FILING DATE: 1998-06-11
; PRIOR APPLICATION NUMBER: 60/089105
; PRIOR FILING DATE: 1998-06-12
; PRIOR APPLICATION NUMBER: 60/089440
; PRIOR FILING DATE: 1998-06-16
; PRIOR APPLICATION NUMBER: 60/089512
; PRIOR FILING DATE: 1998-06-16
; PRIOR APPLICATION NUMBER: 60/089514
; PRIOR FILING DATE: 1998-06-16
; PRIOR APPLICATION NUMBER: 60/089532
; PRIOR FILING DATE: 1998-06-17
; PRIOR APPLICATION NUMBER: 60/089538
; PRIOR FILING DATE: 1998-06-17
; PRIOR APPLICATION NUMBER: 60/089598
; PRIOR FILING DATE: 1998-06-17
; PRIOR APPLICATION NUMBER: 60/089599
; PRIOR FILING DATE: 1998-06-17
; PRIOR APPLICATION NUMBER: 60/089600
; PRIOR FILING DATE: 1998-06-17
; PRIOR APPLICATION NUMBER: 60/089653
; PRIOR FILING DATE: 1998-06-17
; PRIOR APPLICATION NUMBER: 60/089801
; PRIOR FILING DATE: 1998-06-18
; PRIOR APPLICATION NUMBER: 60/089907
; PRIOR FILING DATE: 1998-06-18
; PRIOR APPLICATION NUMBER: 60/089908
; PRIOR FILING DATE: 1998-06-18
; PRIOR APPLICATION NUMBER: 60/089947
; PRIOR FILING DATE: 1998-06-19
; PRIOR APPLICATION NUMBER: 60/089948
; PRIOR FILING DATE: 1998-06-19
; PRIOR APPLICATION NUMBER: 60/089952
; PRIOR FILING DATE: 1998-06-19
; PRIOR APPLICATION NUMBER: 60/090246
; PRIOR FILING DATE: 1998-06-22
; PRIOR APPLICATION NUMBER: 60/090252
; PRIOR FILING DATE: 1998-06-22
; PRIOR APPLICATION NUMBER: 60/090254
; PRIOR FILING DATE: 1998-06-22
; PRIOR APPLICATION NUMBER: 60/090349
; PRIOR FILING DATE: 1998-06-23
; PRIOR APPLICATION NUMBER: 60/090355
; PRIOR FILING DATE: 1998-06-23
; PRIOR APPLICATION NUMBER: 60/090429
; PRIOR FILING DATE: 1998-06-24
; PRIOR APPLICATION NUMBER: 60/090431
; PRIOR FILING DATE: 1998-06-24

; PRIOR APPLICATION NUMBER: 60/090435
; PRIOR FILING DATE: 1998-06-24
; PRIOR APPLICATION NUMBER: 60/090444
; PRIOR FILING DATE: 1998-06-24
; PRIOR APPLICATION NUMBER: 60/090445
; PRIOR FILING DATE: 1998-06-24
; PRIOR APPLICATION NUMBER: 60/090472
; PRIOR FILING DATE: 1998-06-24
; PRIOR APPLICATION NUMBER: 60/090535
; PRIOR FILING DATE: 1998-06-24
; PRIOR APPLICATION NUMBER: 60/090540
; PRIOR FILING DATE: 1998-06-24
; PRIOR APPLICATION NUMBER: 60/090542
; PRIOR FILING DATE: 1998-06-24
; PRIOR APPLICATION NUMBER: 60/090557
; PRIOR FILING DATE: 1998-06-24
; PRIOR APPLICATION NUMBER: 60/090676
; PRIOR FILING DATE: 1998-06-25
; PRIOR APPLICATION NUMBER: 60/090678
; PRIOR FILING DATE: 1998-06-25
; PRIOR APPLICATION NUMBER: 60/090690
; PRIOR FILING DATE: 1998-06-25
; PRIOR APPLICATION NUMBER: 60/090694
; PRIOR FILING DATE: 1998-06-25
; PRIOR APPLICATION NUMBER: 60/090695
; PRIOR FILING DATE: 1998-06-25
; PRIOR APPLICATION NUMBER: 60/090696
; PRIOR FILING DATE: 1998-06-25
; PRIOR APPLICATION NUMBER: 60/090862
; PRIOR FILING DATE: 1998-06-26
; PRIOR APPLICATION NUMBER: 60/090863
; PRIOR FILING DATE: 1998-06-26
; PRIOR APPLICATION NUMBER: 60/091360
; PRIOR FILING DATE: 1998-07-01
; PRIOR APPLICATION NUMBER: 60/091478
; PRIOR FILING DATE: 1998-07-02
; PRIOR APPLICATION NUMBER: 60/091544
; PRIOR FILING DATE: 1998-07-01
; PRIOR APPLICATION NUMBER: 60/091519
; PRIOR FILING DATE: 1998-07-02
; PRIOR APPLICATION NUMBER: 60/091626
; PRIOR FILING DATE: 1998-07-02
; PRIOR APPLICATION NUMBER: 60/091633
; PRIOR FILING DATE: 1998-07-02
; PRIOR APPLICATION NUMBER: 60/091978
; PRIOR FILING DATE: 1998-07-07
; PRIOR APPLICATION NUMBER: 60/091982
; PRIOR FILING DATE: 1998-07-07
; PRIOR APPLICATION NUMBER: 60/092182
; PRIOR FILING DATE: 1998-07-09

Query Match 1.6%; Score 10; DB 2; Length 255;
Best Local Similarity 100.0%; Pred. No. 0.44;
Matches 10; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

```
Qy        381 ENEEEEEEEE 390
              ||||||||||
Db         46 ENEEEEEEEE 55
```

Search completed: February 28, 2006, 09:11:26
Job time : 48 secs

OM protein - protein search, using sw model

Run on: February 28, 2006, 09:10:46 ; Search time 166 Seconds
(without alignments)
1608.391 Million cell updates/sec

Title: US-10-717-665A-44
Perfect score: 639
Sequence: 1 MPQPSVSGMDPPFGDAFRSH.....TAEGNPTGGLVGLRIPTSKV 639

Scoring table: OLIGO
Gapop 60.0 , Gapext 60.0

Searched: 1867569 seqs, 417829326 residues

Word size : 0

Total number of hits satisfying chosen parameters: 1867569

Minimum DB seq length: 0
Maximum DB seq length: 2000000000

Post-processing: Listing first 45 summaries

Database : Published_Applications_AA_Main:*
1: /cgn2_6/ptodata/1/pubpaa/US07_PUBCOMB.pep:*
2: /cgn2_6/ptodata/1/pubpaa/US08_PUBCOMB.pep:*
3: /cgn2_6/ptodata/1/pubpaa/US09_PUBCOMB.pep:*
4: /cgn2_6/ptodata/1/pubpaa/US10A_PUBCOMB.pep:*
5: /cgn2_6/ptodata/1/pubpaa/US10B_PUBCOMB.pep:*
6: /cgn2_6/ptodata/1/pubpaa/US11_PUBCOMB.pep:*

Pred. No. is the number of results predicted by chance to have a score greater than or equal to the score of the result being printed, and is derived by analysis of the total score distribution.

SUMMARIES

Result No.	Score	% Query Match	Length	DB	ID	Description
1	639	100.0	639	5	US-10-717-665-44	Sequence 44, Appl
2	83	13.0	91	5	US-10-717-665-75	Sequence 75, Appl
3	40	6.3	160	5	US-10-450-763-59634	Sequence 59634, A
4	11	1.7	830	4	US-10-211-962-35	Sequence 35, Appl
5	11	1.7	1483	5	US-10-723-860-3730	Sequence 3730, Ap
6	11	1.7	1525	3	US-09-839-479-68	Sequence 68, Appl
7	11	1.7	1525	4	US-10-376-537-69	Sequence 69, Appl
8	11	1.7	1525	4	US-10-702-148-68	Sequence 68, Appl
9	11	1.7	1527	3	US-09-839-479-27	Sequence 27, Appl
10	11	1.7	1527	4	US-10-376-537-27	Sequence 27, Appl
11	11	1.7	1527	4	US-10-702-148-27	Sequence 27, Appl
12	11	1.7	1531	3	US-09-839-479-29	Sequence 29, Appl
13	11	1.7	1531	4	US-10-376-537-29	Sequence 29, Appl
14	11	1.7	1531	4	US-10-702-148-29	Sequence 29, Appl
15	10	1.6	102	4	US-10-425-115-204867	Sequence 204867,
16	10	1.6	236	4	US-10-202-480-4	Sequence 4, Appli
17	10	1.6	255	3	US-09-989-722-287	Sequence 287, App
18	10	1.6	255	3	US-09-989-723-287	Sequence 287, App
19	10	1.6	255	3	US-09-989-279-287	Sequence 287, App
20	10	1.6	255	3	US-09-989-727-287	Sequence 287, App
21	10	1.6	255	3	US-09-989-731-287	Sequence 287, App
22	10	1.6	255	3	US-09-989-732-287	Sequence 287, App
23	10	1.6	255	3	US-09-991-073-287	Sequence 287, App
24	10	1.6	255	3	US-09-990-442-287	Sequence 287, App
25	10	1.6	255	3	US-09-991-163-287	Sequence 287, App
26	10	1.6	255	3	US-09-993-604-287	Sequence 287, App
27	10	1.6	255	3	US-09-990-456-287	Sequence 287, App
28	10	1.6	255	3	US-09-989-721-287	Sequence 287, App
29	10	1.6	255	3	US-09-992-598-287	Sequence 287, App
30	10	1.6	255	3	US-09-989-293A-287	Sequence 287, App
31	10	1.6	255	3	US-09-989-735-287	Sequence 287, App
32	10	1.6	255	3	US-09-990-444-287	Sequence 287, App
33	10	1.6	255	3	US-09-991-181-287	Sequence 287, App
34	10	1.6	255	3	US-09-989-730-287	Sequence 287, App
35	10	1.6	255	3	US-09-990-436-287	Sequence 287, App
36	10	1.6	255	3	US-09-993-687-287	Sequence 287, App
37	10	1.6	255	3	US-09-989-734-287	Sequence 287, App
38	10	1.6	255	3	US-09-997-653-287	Sequence 287, App
39	10	1.6	255	3	US-09-989-724-287	Sequence 287, App
40	10	1.6	255	3	US-09-989-728-287	Sequence 287, App
41	10	1.6	255	3	US-09-990-441-287	Sequence 287, App
42	10	1.6	255	3	US-09-993-667-287	Sequence 287, App
43	10	1.6	255	3	US-09-997-428-287	Sequence 287, App
44	10	1.6	255	3	US-09-997-666-287	Sequence 287, App
45	10	1.6	255	3	US-09-990-438-287	Sequence 287, App

ALIGNMENTS

RESULT 1
US-10-717-665-44
; Sequence 44, Application US/10717665
; Publication No. US20050106579A1
; GENERAL INFORMATION:
; APPLICANT: OriGene Technologies, Inc
; TITLE OF INVENTION: Regulated Angiogenesis Genes and Polypeptides
; FILE REFERENCE: 1U 103 R1
; CURRENT APPLICATION NUMBER: US/10/717,665
; CURRENT FILING DATE: 2003-11-21
; PRIOR APPLICATION NUMBER: US/10/164,595
; PRIOR FILING DATE: 2002-06-10
; NUMBER OF SEQ ID NOS: 80
; SOFTWARE: PatentIn version 3.1
; SEQ ID NO 44
; LENGTH: 639
; TYPE: PRT
; ORGANISM: Homo sapiens
US-10-717-665-44

Query Match 100.0%; Score 639; DB 5; Length 639;
Best Local Similarity 100.0%; Pred. No. 0;
Matches 639; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

```
Qy          1 MPQPSVSGMDPPFGDAFRSHTFSEQTLMSTDLLANSSDPDFMYELDREMNYQQNPRDNFL 60
              ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db          1 MPQPSVSGMDPPFGDAFRSHTFSEQTLMSTDLLANSSDPDFMYELDREMNYQQNPRDNFL 60

Qy         61 SLEDCKDIENLESFTDVLDNEGALTSNWEQWDTYCEDLTKYTKLTSCDIWGTKEVDYLGL 120
              ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db         61 SLEDCKDIENLESFTDVLDNEGALTSNWEQWDTYCEDLTKYTKLTSCDIWGTKEVDYLGL 120

Qy        121 DDFSSPYQDEEVISKTPTLAQLNSEDSQSVSDSLYYPDSLFSVKQNPLPSSFPGKKITSR 180
              ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db        121 DDFSSPYQDEEVISKTPTLAQLNSEDSQSVSDSLYYPDSLFSVKQNPLPSSFPGKKITSR 180

Qy        181 AAAPVCSSKTLQAEVPLSDCVQKASKPPSSTQIMVKTNMYHNEKVNFHVECKDYVKKAKV 240
              ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db        181 AAAPVCSSKTLQAEVPLSDCVQKASKPPSSTQIMVKTNMYHNEKVNFHVECKDYVKKAKV 240

Qy        241 KINPVQQSRPLLSQIHTDAAKENTCYCGAVAKRQEKKGMEPLQGHATPALPFKETQELLL 300
              ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db        241 KINPVQQSRPLLSQIHTDAAKENTCYCGAVAKRQEKKGMEPLQGHATPALPFKETQELLL 300

Qy        301 SPLPQEGPGSLAAGESSSLSASTSVSDSSQKKEEHNYSLFVSDNLGEQPTKCSPEEDEED 360
              ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db        301 SPLPQEGPGSLAAGESSSLSASTSVSDSSQKKEEHNYSLFVSDNLGEQPTKCSPEEDEED 360

Qy        361 EEDVDDEDHDEGFGSEHELSENEEEEEEEEDYEDDKDDDISDTFSEPGYENDSVEDLKEV 420
              ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
```

```
Db        361 EEDVDDEDHDEGFGSEHELSENEEEEEEEEDYEDDKDDDISDTFSEPGYENDSVEDLKEV 420

Qy        421 TSISSRKRGKRRYFWEYSEQLTPSQQERMLRPSEWNRDTLPSNMYQKNGLHHGKYAVKKS 480
              ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db        421 TSISSRKRGKRRYFWEYSEQLTPSQQERMLRPSEWNRDTLPSNMYQKNGLHHGKYAVKKS 480

Qy        481 RRTDVEDLTPNPKKLLQIGNELRKLNKVISDLTPVSELPLTARPRSRKEKNKLAFRACRL 540
              ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db        481 RRTDVEDLTPNPKKLLQIGNELRKLNKVISDLTPVSELPLTARPRSRKEKNKLAFRACRL 540

Qy        541 KKKAQYEANKVKLWGLNTEYDNLLFVINSIKQEIVNRVQNPRDERGPNMGQKLEILIKDT 600
              ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db        541 KKKAQYEANKVKLWGLNTEYDNLLFVINSIKQEIVNRVQNPRDERGPNMGQKLEILIKDT 600

Qy        601 LGLPVAGQTSEFVNQVLEKTAEGNPTGGLVGLRIPTSKV 639
              |||||||||||||||||||||||||||||||||||||||
Db        601 LGLPVAGQTSEFVNQVLEKTAEGNPTGGLVGLRIPTSKV 639


RESULT 2
US-10-717-665-75
; Sequence 75, Application US/10717665
; Publication No. US20050106579A1
; GENERAL INFORMATION:
;  APPLICANT: OriGene Technologies, Inc
;  TITLE OF INVENTION: Regulated Angiogenesis Genes and Polypeptides
;  FILE REFERENCE: 1U 103 R1
;  CURRENT APPLICATION NUMBER: US/10/717,665
;  CURRENT FILING DATE:  2003-11-21
;  PRIOR APPLICATION NUMBER: US/10/164,595
;  PRIOR FILING DATE: 2002-06-10
;  NUMBER OF SEQ ID NOS: 80
;  SOFTWARE: PatentIn version 3.1
; SEQ ID NO 75
;   LENGTH: 91
;   TYPE: PRT
;   ORGANISM: Homo sapiens
US-10-717-665-75

  Query Match             13.0%;  Score 83;  DB 5;  Length 91;
  Best Local Similarity   100.0%;  Pred. No. 9.4e-70;
  Matches   83;  Conservative    0;  Mismatches    0;  Indels    0;  Gaps    0;

Qy        449 MLRPSEWNRDTLPSNMYQKNGLHHGKYAVKKSRRTDVEDLTPNPKKLLQIGNELRKLNKV 508
              |||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db          1 MLRPSEWNRDTLPSNMYQKNGLHHGKYAVKKSRRTDVEDLTPNPKKLLQIGNELRKLNKV 60

Qy        509 ISDLTPVSELPLTARPRSRKEKN 531
              |||||||||||||||||||||||
Db         61 ISDLTPVSELPLTARPRSRKEKN 83


RESULT 3
US-10-450-763-59634
; Sequence 59634, Application US/10450763
; Publication No. US20050196754A1
; GENERAL INFORMATION:
;  APPLICANT: Hyseq, Inc
;  TITLE OF INVENTION: NOVEL NUCLEIC ACIDS AND POLYPEPTIDES
;  FILE REFERENCE: 790CIP3/US
;  CURRENT APPLICATION NUMBER: US/10/450,763
;  CURRENT FILING DATE:  2003-06-11
;  PRIOR APPLICATION NUMBER: PCT/US01/08631
;  PRIOR FILING DATE: 2001-03-30
;  PRIOR APPLICATION NUMBER: 09/540,217
;  PRIOR FILING DATE: 2000-03-31
;  PRIOR APPLICATION NUMBER: 09/649,167
;  PRIOR FILING DATE: 2000-08-23
;  NUMBER OF SEQ ID NOS: 60736
;  SOFTWARE: Custom
; SEQ ID NO 59634
;   LENGTH: 160
;   TYPE: PRT
;   ORGANISM: Homo sapiens
US-10-450-763-59634

  Query Match             6.3%;  Score 40;  DB 5;  Length 160;
  Best Local Similarity   100.0%;  Pred. No. 5.7e-29;
  Matches   40;  Conservative    0;  Mismatches    0;  Indels    0;  Gaps    0;

Qy        562 NLLFVINSIKQEIVNRVQNPRDERGPNMGQKLEILIKDTL 601
              |||||||||||||||||||||||||||||||||||||||||
Db         29 NLLFVINSIKQEIVNRVQNPRDERGPNMGQKLEILIKDTL 68


RESULT 4
US-10-211-962-35
; Sequence 35, Application US/10211962
; Publication No. US20030082640A1
; GENERAL INFORMATION:
;  APPLICANT: Herz, Joachim
;  APPLICANT:  Gotthardt, Michael
;  TITLE OF INVENTION: LDL Receptor Signaling Pathways
;  FILE REFERENCE: UTSW0708
;  CURRENT APPLICATION NUMBER: US/10/211,962
;  CURRENT FILING DATE:  2002-08-01
;  PRIOR APPLICATION NUMBER: US/09/562,737
;  PRIOR FILING DATE: 2000-05-01
;  NUMBER OF SEQ ID NOS: 132
;  SOFTWARE: PatentIn Ver. 2.1
; SEQ ID NO 35
;   LENGTH: 830
;   TYPE: PRT
;   ORGANISM: Artificial Sequence
;   FEATURE:
;   OTHER INFORMATION: Description of Artificial Sequence: Synthetic
;   OTHER INFORMATION: Sequence
US-10-211-962-35

  Query Match             1.7%;  Score 11;  DB 4;  Length 830;
  Best Local Similarity   100.0%;  Pred. No. 0.55;
  Matches   11;  Conservative    0;  Mismatches    0;  Indels    0;  Gaps    0;

Qy        381 ENEEEEEEEED 391
              ||||||||||||
Db         94 ENEEEEEEEED 104


RESULT 5
US-10-723-860-3730
; Sequence 3730, Application US/10723860
; Publication No. US20040253606A1
; GENERAL INFORMATION:
;  APPLICANT: Aziz, Natasha
;  APPLICANT:  Ginsburg, Wendy M.
;  APPLICANT:  Zlotnik, Albert
;  TITLE OF INVENTION: Methods of Diagnosis of Soft Tissue Sarcoma, Compositions &
;  TITLE OF INVENTION:  Methods for Screening for Soft Tissue Sarcoma Modulators
;  FILE REFERENCE: 05882.0193.NPUS01
;  CURRENT APPLICATION NUMBER: US/10/723,860
;  CURRENT FILING DATE:  2003-11-26
;  PRIOR APPLICATION NUMBER: 60/429,739
;  PRIOR FILING DATE: 2002-11-26
;  NUMBER OF SEQ ID NOS: 8393
;  SOFTWARE: PatentIn version 3.2
; SEQ ID NO 3730
;   LENGTH: 1483
;   TYPE: PRT
;   ORGANISM: Homo sapiens
US-10-723-860-3730

  Query Match             1.7%;  Score 11;  DB 5;  Length 1483;
  Best Local Similarity   100.0%;  Pred. No. 0.91;
  Matches   11;  Conservative    0;  Mismatches    0;  Indels    0;  Gaps    0;

Qy        383 EEEEEEEEDYE 393
```

```
                  ||||||||||
Db       1266 EEEEEEEEDYE 1276


RESULT 6
US-09-839-479-68
; Sequence 68, Application US/09839479
; Publication No. US20020039779A1
; GENERAL INFORMATION:
;  APPLICANT: Jones, Michael H.
;  TITLE OF INVENTION: TRANSCRIPTIONAL REGULATOR
;  FILE REFERENCE: 06501-042002
;  CURRENT APPLICATION NUMBER: US/09/839,479
;  CURRENT FILING DATE:  2001-04-20
;  PRIOR APPLICATION NUMBER: US 09/418,710
;  PRIOR FILING DATE: 1999-10-15
;  PRIOR APPLICATION NUMBER: PCT/JP98/01783
;  PRIOR FILING DATE: 1998-04-17
;  PRIOR APPLICATION NUMBER: JP 9/310027
;  PRIOR FILING DATE: 1997-10-24
;  PRIOR APPLICATION NUMBER: JP 9/116570
;  PRIOR FILING DATE: 1997-04-18
;  NUMBER OF SEQ ID NOS: 72
;  SOFTWARE: FastSEQ for Windows Version 4.0
; SEQ ID NO 68
;   LENGTH: 1525
;   TYPE: PRT
;   ORGANISM: Homo sapiens
US-09-839-479-68

  Query Match             1.7%;  Score 11;  DB 3;  Length 1525;
  Best Local Similarity   100.0%;  Pred. No. 0.94;
  Matches   11;  Conservative    0;  Mismatches    0;  Indels    0;  Gaps    0;

Qy        383 EEEEEEEEDYE 393
              ||||||||||
Db       1264 EEEEEEEEDYE 1274


RESULT 7
US-10-376-537-69
; Sequence 69, Application US/10376537
; Publication No. US20030224405A1
; GENERAL INFORMATION:
;  APPLICANT: Jones, Michael H.
;  TITLE OF INVENTION: TRANSCRIPTIONAL REGULATOR
;  FILE REFERENCE: 06501-042001
;  CURRENT APPLICATION NUMBER: US/10/376,537
;  CURRENT FILING DATE:  2003-02-28
;  PRIOR APPLICATION NUMBER: US/09/418,710
;  PRIOR FILING DATE: 1999-10-15
;  PRIOR APPLICATION NUMBER: PCT/JP98/01783
;  PRIOR FILING DATE: 1998-04-17
;  PRIOR APPLICATION NUMBER: JP 9/310027
;  PRIOR FILING DATE: 1997-10-24
;  PRIOR APPLICATION NUMBER: JP 9/116570
;  PRIOR FILING DATE: 1997-04-18
;  NUMBER OF SEQ ID NOS: 73
;  SOFTWARE: FastSEQ for Windows Version 4.0
; SEQ ID NO 69
;   LENGTH: 1525
;   TYPE: PRT
;   ORGANISM: Homo sapiens
US-10-376-537-69

  Query Match             1.7%;  Score 11;  DB 4;  Length 1525;
  Best Local Similarity   100.0%;  Pred. No. 0.94;
  Matches   11;  Conservative    0;  Mismatches    0;  Indels    0;  Gaps    0;

Qy        383 EEEEEEEEDYE 393
              ||||||||||
Db       1264 EEEEEEEEDYE 1274


RESULT 8
US-10-702-148-68
; Sequence 68, Application US/10702148
; Publication No. US20040063145A1
; GENERAL INFORMATION:
;  APPLICANT: Jones, Michael H.
;  TITLE OF INVENTION: TRANSCRIPTIONAL REGULATOR
;  FILE REFERENCE: 06501-042002
;  CURRENT APPLICATION NUMBER: US/10/702,148
;  CURRENT FILING DATE:  2003-11-05
;  PRIOR APPLICATION NUMBER: PRIOR  APPLICATION NUMBER: US/09/839,479
;  PRIOR FILING DATE: 2001-04-20
;  PRIOR APPLICATION NUMBER: US 09/418,710
;  PRIOR FILING DATE: 1999-10-15
;  PRIOR APPLICATION NUMBER: PCT/JP98/01783
;  PRIOR FILING DATE: 1998-04-17
;  PRIOR APPLICATION NUMBER: JP 9/310027
;  PRIOR FILING DATE: 1997-10-24
;  PRIOR APPLICATION NUMBER: JP 9/116570
;  PRIOR FILING DATE: 1997-04-18
;  NUMBER OF SEQ ID NOS: 72
;  SOFTWARE: FastSEQ for Windows Version 4.0
; SEQ ID NO 68
;   LENGTH: 1525
;   TYPE: PRT
;   ORGANISM: Homo sapiens
US-10-702-148-68

  Query Match             1.7%;  Score 11;  DB 4;  Length 1525;
  Best Local Similarity   100.0%;  Pred. No. 0.94;
  Matches   11;  Conservative    0;  Mismatches    0;  Indels    0;  Gaps    0;

Qy        383 EEEEEEEEDYE 393
              ||||||||||
Db       1264 EEEEEEEEDYE 1274


RESULT 9
US-09-839-479-27
; Sequence 27, Application US/09839479
; Publication No. US20020039779A1
; GENERAL INFORMATION:
;  APPLICANT: Jones, Michael H.
;  TITLE OF INVENTION: TRANSCRIPTIONAL REGULATOR
;  FILE REFERENCE: 06501-042002
;  CURRENT APPLICATION NUMBER: US/09/839,479
;  CURRENT FILING DATE:  2001-04-20
;  PRIOR APPLICATION NUMBER: US 09/418,710
;  PRIOR FILING DATE: 1999-10-15
;  PRIOR APPLICATION NUMBER: PCT/JP98/01783
;  PRIOR FILING DATE: 1998-04-17
;  PRIOR APPLICATION NUMBER: JP 9/310027
;  PRIOR FILING DATE: 1997-10-24
;  PRIOR APPLICATION NUMBER: JP 9/116570
;  PRIOR FILING DATE: 1997-04-18
;  NUMBER OF SEQ ID NOS: 72
;  SOFTWARE: FastSEQ for Windows Version 4.0
; SEQ ID NO 27
;   LENGTH: 1527
;   TYPE: PRT
;   ORGANISM: Homo sapiens
US-09-839-479-27

  Query Match             1.7%;  Score 11;  DB 3;  Length 1527;
  Best Local Similarity   100.0%;  Pred. No. 0.94;
  Matches   11;  Conservative    0;  Mismatches    0;  Indels    0;  Gaps    0;

Qy        383 EEEEEEEEDYE 393
              ||||||||||
Db       1266 EEEEEEEEDYE 1276
```

```
RESULT 10
US-10-376-537-27
; Sequence 27, Application US/10376537
; Publication No. US20030224405A1
; GENERAL INFORMATION:
;  APPLICANT: Jones, Michael H.
;  TITLE OF INVENTION: TRANSCRIPTIONAL REGULATOR
;  FILE REFERENCE: 06501-042001
;  CURRENT APPLICATION NUMBER: US/10/376,537
;  CURRENT FILING DATE:  2003-02-28
;  PRIOR APPLICATION NUMBER: US/09/418,710
;  PRIOR FILING DATE: 1999-10-15
;  PRIOR APPLICATION NUMBER: PCT/JP98/01783
;  PRIOR FILING DATE: 1998-04-17
;  PRIOR APPLICATION NUMBER: JP 9/310027
;  PRIOR FILING DATE: 1997-10-24
;  PRIOR APPLICATION NUMBER: JP 9/116570
;  PRIOR FILING DATE: 1997-04-18
;  NUMBER OF SEQ ID NOS: 73
;  SOFTWARE: FastSEQ for Windows Version 4.0
; SEQ ID NO 27
;   LENGTH: 1527
;   TYPE: PRT
;   ORGANISM: Homo sapiens
US-10-376-537-27

  Query Match             1.7%;  Score 11;  DB 4;  Length 1527;
  Best Local Similarity   100.0%;  Pred. No. 0.94;
  Matches   11;  Conservative    0;  Mismatches    0;  Indels    0;  Gaps    0;

Qy        383 EEEEEEEEDYE 393
              ||||||||||| |
Db       1266 EEEEEEEEDYE 1276


RESULT 11
US-10-702-148-27
; Sequence 27, Application US/10702148
; Publication No. US20040063145A1
; GENERAL INFORMATION:
;  APPLICANT: Jones, Michael H.
;  TITLE OF INVENTION: TRANSCRIPTIONAL REGULATOR
;  FILE REFERENCE: 06501-042002
;  CURRENT APPLICATION NUMBER: US/10/702,148
;  CURRENT FILING DATE:  2003-11-05
;  PRIOR APPLICATION NUMBER: PRIOR  APPLICATION NUMBER: US/09/839,479
;  PRIOR FILING DATE: 2001-04-20
;  PRIOR APPLICATION NUMBER: US 09/418,710
;  PRIOR FILING DATE: 1999-10-15
;  PRIOR APPLICATION NUMBER: PCT/JP98/01783
;  PRIOR FILING DATE: 1998-04-17
;  PRIOR APPLICATION NUMBER: JP 9/310027
;  PRIOR FILING DATE: 1997-10-24
;  PRIOR APPLICATION NUMBER: JP 9/116570
;  PRIOR FILING DATE: 1997-04-18
;  NUMBER OF SEQ ID NOS: 72
;  SOFTWARE: FastSEQ for Windows Version 4.0
; SEQ ID NO 27
;   LENGTH: 1527
;   TYPE: PRT
;   ORGANISM: Homo sapiens
US-10-702-148-27

  Query Match             1.7%;  Score 11;  DB 4;  Length 1527;
  Best Local Similarity   100.0%;  Pred. No. 0.94;
  Matches   11;  Conservative    0;  Mismatches    0;  Indels    0;  Gaps    0;

Qy        383 EEEEEEEEDYE 393
              ||||||||||| |
Db       1266 EEEEEEEEDYE 1276


RESULT 12
US-09-839-479-29
; Sequence 29, Application US/09839479
; Publication No. US20020039779A1
; GENERAL INFORMATION:
;  APPLICANT: Jones, Michael H.
;  TITLE OF INVENTION: TRANSCRIPTIONAL REGULATOR
;  FILE REFERENCE: 06501-042002
;  CURRENT APPLICATION NUMBER: US/09/839,479
;  CURRENT FILING DATE:  2001-04-20
;  PRIOR APPLICATION NUMBER: US 09/418,710
;  PRIOR FILING DATE: 1999-10-15
;  PRIOR APPLICATION NUMBER: PCT/JP98/01783
;  PRIOR FILING DATE: 1998-04-17
;  PRIOR APPLICATION NUMBER: JP 9/310027
;  PRIOR FILING DATE: 1997-10-24
;  PRIOR APPLICATION NUMBER: JP 9/116570
;  PRIOR FILING DATE: 1997-04-18
;  NUMBER OF SEQ ID NOS: 72
;  SOFTWARE: FastSEQ for Windows Version 4.0
; SEQ ID NO 29
;   LENGTH: 1531
;   TYPE: PRT
;   ORGANISM: Homo sapiens
US-09-839-479-29

  Query Match             1.7%;  Score 11;  DB 3;  Length 1531;
  Best Local Similarity   100.0%;  Pred. No. 0.94;
  Matches   11;  Conservative    0;  Mismatches    0;  Indels    0;  Gaps    0;

Qy        383 EEEEEEEEDYE 393
              ||||||||||| |
Db       1270 EEEEEEEEDYE 1280


RESULT 13
US-10-376-537-29
; Sequence 29, Application US/10376537
; Publication No. US20030224405A1
; GENERAL INFORMATION:
;  APPLICANT: Jones, Michael H.
;  TITLE OF INVENTION: TRANSCRIPTIONAL REGULATOR
;  FILE REFERENCE: 06501-042001
;  CURRENT APPLICATION NUMBER: US/10/376,537
;  CURRENT FILING DATE:  2003-02-28
;  PRIOR APPLICATION NUMBER: US/09/418,710
;  PRIOR FILING DATE: 1999-10-15
;  PRIOR APPLICATION NUMBER: PCT/JP98/01783
;  PRIOR FILING DATE: 1998-04-17
;  PRIOR APPLICATION NUMBER: JP 9/310027
;  PRIOR FILING DATE: 1997-10-24
;  PRIOR APPLICATION NUMBER: JP 9/116570
;  PRIOR FILING DATE: 1997-04-18
;  NUMBER OF SEQ ID NOS: 73
;  SOFTWARE: FastSEQ for Windows Version 4.0
; SEQ ID NO 29
;   LENGTH: 1531
;   TYPE: PRT
;   ORGANISM: Homo sapiens
US-10-376-537-29

  Query Match             1.7%;  Score 11;  DB 4;  Length 1531;
  Best Local Similarity   100.0%;  Pred. No. 0.94;
  Matches   11;  Conservative    0;  Mismatches    0;  Indels    0;  Gaps    0;

Qy        383 EEEEEEEEDYE 393
              ||||||||||| |
Db       1270 EEEEEEEEDYE 1280


RESULT 14
US-10-702-148-29
; Sequence 29, Application US/10702148
; Publication No. US20040063145A1
```

```
; GENERAL INFORMATION:
;  APPLICANT: Jones, Michael H.
;  TITLE OF INVENTION: TRANSCRIPTIONAL REGULATOR
;  FILE REFERENCE: 06501-042002
;  CURRENT APPLICATION NUMBER: US/10/702,148
;  CURRENT FILING DATE:  2003-11-05
;  PRIOR APPLICATION NUMBER: PRIOR  APPLICATION NUMBER: US/09/839,479
;  PRIOR FILING DATE: 2001-04-20
;  PRIOR APPLICATION NUMBER: US 09/418,710
;  PRIOR FILING DATE: 1999-10-15
;  PRIOR APPLICATION NUMBER: PCT/JP98/01783
;  PRIOR FILING DATE: 1998-04-17
;  PRIOR APPLICATION NUMBER: JP 9/310027
;  PRIOR FILING DATE: 1997-10-24
;  PRIOR APPLICATION NUMBER: JP 9/116570
;  PRIOR FILING DATE: 1997-04-18
;  NUMBER OF SEQ ID NOS: 72
;  SOFTWARE: FastSEQ for Windows Version 4.0
; SEQ ID NO 29
;   LENGTH: 1531
;   TYPE: PRT
;   ORGANISM: Homo sapiens
US-10-702-148-29

  Query Match             1.7%;  Score 11;  DB 4;  Length 1531;
  Best Local Similarity   100.0%;  Pred. No. 0.94;
  Matches   11;  Conservative    0;  Mismatches    0;  Indels    0;  Gaps    0;

Qy        383 EEEEEEEEDYE 393
              ||||||||||
Db       1270 EEEEEEEEDYE 1280


RESULT 15
US-10-425-115-204867
; Sequence 204867, Application US/10425115
; Publication No. US20040214272A1
; GENERAL INFORMATION:
;  APPLICANT: La Rosa, Thomas J.
;  APPLICANT:  Kovalic, David K.
;  APPLICANT:  Zhou, Yihua
;  APPLICANT:  Cao, Yongwei
;  TITLE OF INVENTION: Nucleic Acid Molecules and Other Molecules Associated With
;  TITLE OF INVENTION:  Plants
;  FILE REFERENCE: 38-21(53222)B
;  CURRENT APPLICATION NUMBER: US/10/425,115
;  CURRENT FILING DATE:  2003-04-28
;  NUMBER OF SEQ ID NOS: 369326
; SEQ ID NO 204867
;   LENGTH: 102
;   TYPE: PRT
;   ORGANISM: Zea mays
;   FEATURE:
;   OTHER INFORMATION: Clone ID: MRT4577_118428C.1.pep
US-10-425-115-204867

  Query Match             1.6%;  Score 10;  DB 4;  Length 102;
  Best Local Similarity   100.0%;  Pred. No. 0.76;
  Matches   10;  Conservative    0;  Mismatches    0;  Indels    0;  Gaps    0;

Qy        381 ENEEEEEEEE 390
              ||||||||||
Db         75 ENEEEEEEEE 84


Search completed: February 28, 2006, 09:14:18
Job time : 167 secs
```

GenCore version 5.1.7

OM protein - protein search, using sw model

Run on: February 28, 2006, 09:14:31 ; Search time 19 Seconds
(without alignments)
500.682 Million cell updates/sec

Title: US-10-717-665A-44
Perfect score: 639
Sequence: 1 MPQPSVSGMDPPFGDAFRSH.....TAEGNPTGGLVGLRIPTSKV 639

Scoring table: OLIGO
Gapop 60.0 , Gapext 60.0

Searched: 117670 seqs, 14887254 residues

Word size : 0

Total number of hits satisfying chosen parameters: 117670

Minimum DB seq length: 0
Maximum DB seq length: 2000000000

Post-processing: Listing first 45 summaries

Database : Published_Applications_AA_New:*
1: /cgn2_6/ptodata/2/pubpaa/US08_NEW_PUB.pep:*
2: /cgn2_6/ptodata/2/pubpaa/US06_NEW_PUB.pep:*
3: /cgn2_6/ptodata/2/pubpaa/US07_NEW_PUB.pep:*
4: /cgn2_6/ptodata/2/pubpaa/PCT_NEW_PUB.pep:*
5: /cgn2_6/ptodata/2/pubpaa/US09_NEW_PUB.pep:*
6: /cgn2_6/ptodata/2/pubpaa/US10_NEW_PUB.pep:*
7: /cgn2_6/ptodata/2/pubpaa/US11_NEW_PUB.pep:*
8: /cgn2_6/ptodata/2/pubpaa/US60_NEW_PUB.pep:*

Pred. No. is the number of results predicted by chance to have a score greater than or equal to the score of the result being printed, and is derived by analysis of the total score distribution.

SUMMARIES

Result No.	Score	Query Match	Length	DB	ID	Description
1	10	1.6	593	7	US-11-040-488-2	Sequence 2, Appli
2	9	1.4	215	6	US-10-821-234-1443	Sequence 1443, Ap
3	9	1.4	221	7	US-11-067-425A-74	Sequence 74, Appl
4	9	1.4	229	7	US-11-063-343-34	Sequence 34, Appl
5	9	1.4	268	7	US-11-072-512-3158	Sequence 3158, Ap
6	9	1.4	414	6	US-10-821-234-1170	Sequence 1170, Ap
7	9	1.4	427	6	US-10-714-887-122	Sequence 122, App
8	9	1.4	547	7	US-11-072-512-3096	Sequence 3096, Ap
9	9	1.4	706	7	US-11-072-512-3843	Sequence 3843, Ap
10	9	1.4	732	7	US-11-078-189-14	Sequence 14, Appl
11	9	1.4	1178	7	US-11-044-899-29	Sequence 29, Appl
12	9	1.4	1229	7	US-11-054-281-107	Sequence 107, App
13	9	1.4	2161	7	US-11-126-313-31	Sequence 31, Appl
14	9	1.4	4868	7	US-11-044-111-24	Sequence 24, Appl
15	9	1.4	8746	7	US-11-098-686-10232	Sequence 10232, A
16	8	1.3	91	6	US-10-821-234-1345	Sequence 1345, Ap
17	8	1.3	144	6	US-10-821-234-1254	Sequence 1254, Ap
18	8	1.3	215	7	US-11-067-425A-71	Sequence 71, Appl
19	8	1.3	281	6	US-10-883-512-90	Sequence 90, Appl
20	8	1.3	286	7	US-11-072-512-2850	Sequence 2850, Ap
21	8	1.3	292	7	US-11-124-367A-308	Sequence 308, App
22	8	1.3	303	6	US-10-467-962B-16	Sequence 16, Appl
23	8	1.3	303	6	US-10-467-962B-45	Sequence 45, Appl
24	8	1.3	325	7	US-11-063-343-29	Sequence 29, Appl
25	8	1.3	344	6	US-10-821-234-911	Sequence 911, App
26	8	1.3	354	7	US-11-189-817-2	Sequence 2, Appli
27	8	1.3	363	7	US-11-124-367A-309	Sequence 309, App
28	8	1.3	417	6	US-10-821-234-1536	Sequence 1536, Ap
29	8	1.3	443	7	US-11-054-385-4	Sequence 4, Appli
30	8	1.3	454	6	US-10-509-773-8	Sequence 8, Appli
31	8	1.3	457	6	US-10-982-545-8	Sequence 8, Appli
32	8	1.3	457	6	US-10-982-545-13	Sequence 13, Appl
33	8	1.3	477	6	US-10-467-657-5516	Sequence 5516, Ap
34	8	1.3	495	6	US-10-508-263-20	Sequence 20, Appl
35	8	1.3	526	7	US-11-236-198-19	Sequence 19, Appl
36	8	1.3	565	7	US-11-080-991-100	Sequence 100, App
37	8	1.3	575	7	US-11-236-198-18	Sequence 18, Appl
38	8	1.3	578	6	US-10-821-234-1039	Sequence 1039, Ap
39	8	1.3	580	7	US-11-072-512-2103	Sequence 2103, Ap
40	8	1.3	585	6	US-10-878-556A-42	Sequence 42, Appl
41	8	1.3	594	7	US-11-236-198-17	Sequence 17, Appl
42	8	1.3	605	6	US-10-131-826A-160	Sequence 160, App
43	8	1.3	634	7	US-11-124-367A-285	Sequence 285, App
44	8	1.3	636	7	US-11-072-512-3666	Sequence 3666, Ap
45	8	1.3	642	6	US-10-131-826A-370	Sequence 370, App

ALIGNMENTS

RESULT 1
US-11-040-488-2
; Sequence 2, Application US/11040488
; Publication No. US20050271651A1
; GENERAL INFORMATION:
; APPLICANT: WEBB, CAROL
; TITLE OF INVENTION: INHIBITION OF BRIGHT FUNCTION AS A TREATMENT FOR
; TITLE OF INVENTION: EXCESSIVE IMMUNOGLOBULIN PRODUCTION
; FILE REFERENCE: OMRF:023US
; CURRENT APPLICATION NUMBER: US/11/040,488
; CURRENT FILING DATE: 2005-01-21
; PRIOR APPLICATION NUMBER: 60/538,866
; PRIOR FILING DATE: 2004-01-23
; NUMBER OF SEQ ID NOS: 26
; SOFTWARE: PatentIn Ver. 2.1
; SEQ ID NO 2
; LENGTH: 593
; TYPE: PRT
; ORGANISM: Homo sapiens
US-11-040-488-2

Query Match 1.6%; Score 10; DB 7; Length 593;
Best Local Similarity 100.0%; Pred. No. 0.089;
Matches 10; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

Qy 385 EEEEEEDYED 394
||||||||||
Db 140 EEEEEEDYED 149

RESULT 2
US-10-821-234-1443
; Sequence 1443, Application US/10821234
; Publication No. US20050255114A1
; GENERAL INFORMATION:
; APPLICANT: Labat, Ivan
; APPLICANT: Stache-Crain, Birgit
; APPLICANT: Andarmani, Susan
; APPLICANT: Tang, Y. Tom
; TITLE OF INVENTION: Methods for Diagnosis and Treatment of Preeclampsia
; FILE REFERENCE: 821A
; CURRENT APPLICATION NUMBER: US/10/821,234
; CURRENT FILING DATE: 2004-04-07
; PRIOR APPLICATION NUMBER: US 60/462,047
; PRIOR FILING DATE: 2003-04-07
; NUMBER OF SEQ ID NOS: 1704
; SOFTWARE: pt_SEQ_genes Version 1.0
; SEQ ID NO 1443

; LENGTH: 215
; TYPE: PRT
; ORGANISM: Homo sapiens
US-10-821-234-1443

  Query Match 1.4%; Score 9; DB 6; Length 215;
  Best Local Similarity 100.0%; Pred. No. 0.34;
  Matches 9; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

Qy 355 EEDEEDEED 363
        |||||||||
Db 188 EEDEEDEED 196

RESULT 3
US-11-067-425A-74
; Sequence 74, Application US/11067425A
; Publication No. US20050278809A1
; GENERAL INFORMATION:
; APPLICANT: Hannoufa, Abdelali
; APPLICANT: Lydiate, Derek J.
; APPLICANT: Gao, Ming-Jun
; TITLE OF INVENTION: REGULATION OF GENE EXPRESSION USING CHROMATIN REMODELLING FACTORS
; FILE REFERENCE: 270.78USI1
; CURRENT APPLICATION NUMBER: US/11/067,425A
; CURRENT FILING DATE: 2005-02-22
; PRIOR APPLICATION NUMBER: US 10/516,753
; PRIOR FILING DATE: 2004-12-03
; PRIOR APPLICATION NUMBER: PCT/CA03/00822
; PRIOR FILING DATE: 2003-06-06
; PRIOR APPLICATION NUMBER: US 60/387,088
; PRIOR FILING DATE: 2002-06-06
; NUMBER OF SEQ ID NOS: 108
; SOFTWARE: PatentIn version 3.3
; SEQ ID NO 74
; LENGTH: 221
; TYPE: PRT
; ORGANISM: Arabidopsis
US-11-067-425A-74

  Query Match 1.4%; Score 9; DB 7; Length 221;
  Best Local Similarity 100.0%; Pred. No. 0.35;
  Matches 9; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

Qy 380 SENEEEEEE 388
        |||||||||
Db   72 SENEEEEEE 80

RESULT 4
US-11-063-343-34
; Sequence 34, Application US/11063343
; Publication No. US20050272061A1
; GENERAL INFORMATION:
; APPLICANT: Petroziello, Joseph M.
; APPLICANT: Carter, Paul
; TITLE OF INVENTION: Expression Profiling in Non-Small Cell
; TITLE OF INVENTION: Lung Cancer
; FILE REFERENCE: 2681-1-003N
; CURRENT APPLICATION NUMBER: US/11/063,343
; CURRENT FILING DATE: 2005-02-22
; PRIOR APPLICATION NUMBER: 60/546,019
; PRIOR FILING DATE: 2004-02-19
; NUMBER OF SEQ ID NOS: 50
; SOFTWARE: FastSEQ for Windows Version 4.0
; SEQ ID NO 34
; LENGTH: 229
; TYPE: PRT
; ORGANISM: Homo sapiens
US-11-063-343-34

  Query Match 1.4%; Score 9; DB 7; Length 229;
  Best Local Similarity 100.0%; Pred. No. 0.36;
  Matches 9; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

Qy 359 EDEEDVDDE 367
        |||||||||
Db   85 EDEEDVDDE 93

RESULT 5
US-11-072-512-3158
; Sequence 3158, Application US/11072512
; Publication No. US20060029945A1
; GENERAL INFORMATION:
; APPLICANT: ISOGAI, TAKAO
; APPLICANT: SUGIYAMA, TOMOYASU
; APPLICANT: OTSUKI, TETSUJI
; APPLICANT: WAKAMATSU, AI
; APPLICANT: SATO, HIROYUKI
; APPLICANT: ISHII, SHIZUKO
; APPLICANT: YAMAMOTO, JUN-ICHI
; APPLICANT: ISONO, YUUKO
; APPLICANT: HIO, YURI
; APPLICANT: OTSUKA, KAORU
; APPLICANT: NAGAI, KEIICHI
; APPLICANT: IRIE, RYOTARO
; APPLICANT: TAMECHIKA, ICHIRO
; APPLICANT: SEKI, NAOHIKO
; APPLICANT: YOSHIKAWA, TSUTOMU
; APPLICANT: OTSUKA, MOTOYUKI
; APPLICANT: NAGAHARI, KENJI
; APPLICANT: MASUHO, YASUHIKO
; TITLE OF INVENTION: Novel full length cDNA
; FILE REFERENCE: 084335-0191
; CURRENT APPLICATION NUMBER: US/11/072,512
; CURRENT FILING DATE: 2005-03-07
; PRIOR APPLICATION NUMBER: US 60/350,978
; PRIOR FILING DATE: 2002-01-25
; PRIOR APPLICATION NUMBER: JP 2001-379298
; PRIOR FILING DATE: 2001-11-05
; NUMBER OF SEQ ID NOS: 4096
; SOFTWARE: PatentIn Ver. 2.1
; SEQ ID NO 3158
; LENGTH: 268
; TYPE: PRT
; ORGANISM: Homo sapiens
US-11-072-512-3158

  Query Match 1.4%; Score 9; DB 7; Length 268;
  Best Local Similarity 100.0%; Pred. No. 0.42;
  Matches 9; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

Qy 383 EEEEEEEED 391
        |||||||||
Db 187 EEEEEEEED 195

RESULT 6
US-10-821-234-1170
; Sequence 1170, Application US/10821234
; Publication No. US20050255114A1
; GENERAL INFORMATION:
; APPLICANT: Labat, Ivan
; APPLICANT: Stache-Crain, Birgit
; APPLICANT: Andarmani, Susan
; APPLICANT: Tang, Y. Tom
; TITLE OF INVENTION: Methods for Diagnosis and Treatment of Preeclampsia
; FILE REFERENCE: 821A
; CURRENT APPLICATION NUMBER: US/10/821,234
; CURRENT FILING DATE: 2004-04-07
; PRIOR APPLICATION NUMBER: US 60/462,047
; PRIOR FILING DATE: 2003-04-07
; NUMBER OF SEQ ID NOS: 1704
; SOFTWARE: pt_SEQ_genes Version 1.0
; SEQ ID NO 1170

; LENGTH: 414
; TYPE: PRT
; ORGANISM: Homo sapiens
US-10-821-234-1170

  Query Match 1.4%; Score 9; DB 6; Length 414;
  Best Local Similarity 100.0%; Pred. No. 0.63;
  Matches 9; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

Qy 383 EEEEEEEED 391
        |||||||||
Db 77 EEEEEEEED 85

RESULT 7
US-10-714-887-122
; Sequence 122, Application US/10714887
; Publication No. US20060015972A1
; GENERAL INFORMATION:
; APPLICANT: Mendel Biotechnology, Inc.
; APPLICANT: HEARD, Jacqueline
; APPLICANT: RIECHMANN, Jose Luis
; APPLICANT: CREELMAN, Robert
; APPLICANT: RATCLIFFE, Oliver
; APPLICANT: CANALES, Roger
; APPLICANT: REPETTI, Peter
; APPLICANT: KUMIMOTO, Roderick W
; APPLICANT: GUTTERSON, Neal
; APPLICANT: REUBER, T. Lynne
; APPLICANT: PINEDA, Omaira
; APPLICANT: SHERMAN, Bradley K
; TITLE OF INVENTION: PLANT TRANSCRIPTIONAL REGULATORS OF DROUGHT STRESS
; FILE REFERENCE: MBI0058-CIP
; CURRENT APPLICATION NUMBER: US/10/714,887
; CURRENT FILING DATE: 2003-11-13
; PRIOR APPLICATION NUMBER: 10/412,699
; PRIOR FILING DATE: 2003-04-10
; PRIOR APPLICATION NUMBER: 09/506,720
; PRIOR FILING DATE: 2000-02-17
; PRIOR APPLICATION NUMBER: 60/135,134
; PRIOR FILING DATE: 1999-05-20
; PRIOR APPLICATION NUMBER: 09/394,519
; PRIOR FILING DATE: 1999-09-13
; PRIOR APPLICATION NUMBER: 09/533,392
; PRIOR FILING DATE: 2000-03-22
; PRIOR APPLICATION NUMBER: 09/533,029
; PRIOR FILING DATE: 2000-03-22
; PRIOR APPLICATION NUMBER: 09/532,591
; PRIOR FILING DATE: 2000-03-22
; PRIOR APPLICATION NUMBER: 09/533,030
; PRIOR FILING DATE: 2000-03-22
; PRIOR APPLICATION NUMBER: 60/125,814
; PRIOR FILING DATE: 1999-03-23
; PRIOR APPLICATION NUMBER: 09/713,994
; PRIOR FILING DATE: 2000-11-16
; Remaining Prior Application data removed - See File Wrapper or PALM.
; NUMBER OF SEQ ID NOS: 430
; SOFTWARE: PatentIn version 3.2
; SEQ ID NO 122
; LENGTH: 427
; TYPE: PRT
; ORGANISM: Oryza sativa (japonica cultivar-group)
; FEATURE:
; OTHER INFORMATION: G3690 polypeptide Orthologous to G2999
US-10-714-887-122

  Query Match 1.4%; Score 9; DB 6; Length 427;
  Best Local Similarity 100.0%; Pred. No. 0.65;
  Matches 9; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

Qy 383 EEEEEEEED 391
        |||||||||
Db 12 EEEEEEEED 20

RESULT 8
US-11-072-512-3096
; Sequence 3096, Application US/11072512
; Publication No. US20060029945A1
; GENERAL INFORMATION:
; APPLICANT: ISOGAI, TAKAO
; APPLICANT: SUGIYAMA, TOMOYASU
; APPLICANT: OTSUKI, TETSUJI
; APPLICANT: WAKAMATSU, AI
; APPLICANT: SATO, HIROYUKI
; APPLICANT: ISHII, SHIZUKO
; APPLICANT: YAMAMOTO, JUN-ICHI
; APPLICANT: ISONO, YUUKO
; APPLICANT: HIO, YURI
; APPLICANT: OTSUKA, KAORU
; APPLICANT: NAGAI, KEIICHI
; APPLICANT: IRIE, RYOTARO
; APPLICANT: TAMECHIKA, ICHIRO
; APPLICANT: SEKI, NAOHIKO
; APPLICANT: YOSHIKAWA, TSUTOMU
; APPLICANT: OTSUKA, MOTOYUKI
; APPLICANT: NAGAHARI, KENJI
; APPLICANT: MASUHO, YASUHIKO
; TITLE OF INVENTION: Novel full length cDNA
; FILE REFERENCE: 084335-0191
; CURRENT APPLICATION NUMBER: US/11/072,512
; CURRENT FILING DATE: 2005-03-07
; PRIOR APPLICATION NUMBER: US 60/350,978
; PRIOR FILING DATE: 2002-01-25
; PRIOR APPLICATION NUMBER: JP 2001-379298
; PRIOR FILING DATE: 2001-11-05
; NUMBER OF SEQ ID NOS: 4096
; SOFTWARE: PatentIn Ver. 2.1
; SEQ ID NO 3096
; LENGTH: 547
; TYPE: PRT
; ORGANISM: Homo sapiens
US-11-072-512-3096

  Query Match 1.4%; Score 9; DB 7; Length 547;
  Best Local Similarity 100.0%; Pred. No. 0.82;
  Matches 9; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

Qy 383 EEEEEEEED 391
        |||||||||
Db 56 EEEEEEEED 64

RESULT 9
US-11-072-512-3843
; Sequence 3843, Application US/11072512
; Publication No. US20060029945A1
; GENERAL INFORMATION:
; APPLICANT: ISOGAI, TAKAO
; APPLICANT: SUGIYAMA, TOMOYASU
; APPLICANT: OTSUKI, TETSUJI
; APPLICANT: WAKAMATSU, AI
; APPLICANT: SATO, HIROYUKI
; APPLICANT: ISHII, SHIZUKO
; APPLICANT: YAMAMOTO, JUN-ICHI
; APPLICANT: ISONO, YUUKO
; APPLICANT: HIO, YURI
; APPLICANT: OTSUKA, KAORU
; APPLICANT: NAGAI, KEIICHI
; APPLICANT: IRIE, RYOTARO
; APPLICANT: TAMECHIKA, ICHIRO
; APPLICANT: SEKI, NAOHIKO
; APPLICANT: YOSHIKAWA, TSUTOMU
; APPLICANT: OTSUKA, MOTOYUKI
; APPLICANT: NAGAHARI, KENJI
; APPLICANT: MASUHO, YASUHIKO

```
;  TITLE OF INVENTION: Novel full length cDNA
;  FILE REFERENCE: 084335-0191
;  CURRENT APPLICATION NUMBER: US/11/072,512
;  CURRENT FILING DATE:  2005-03-07
;  PRIOR APPLICATION NUMBER: US 60/350,978
;  PRIOR FILING DATE: 2002-01-25
;  PRIOR APPLICATION NUMBER: JP 2001-379298
;  PRIOR FILING DATE: 2001-11-05
;  NUMBER OF SEQ ID NOS: 4096
;  SOFTWARE: PatentIn Ver. 2.1
; SEQ ID NO 3843
;   LENGTH: 706
;   TYPE: PRT
;   ORGANISM: Homo sapiens
US-11-072-512-3843

  Query Match             1.4%;  Score 9;  DB 7;  Length 706;
  Best Local Similarity   100.0%;  Pred. No. 1;
  Matches    9;  Conservative    0;  Mismatches    0;  Indels    0;  Gaps    0;

Qy        383 EEEEEEEED 391
              |||||||||
Db        253 EEEEEEEED 261


RESULT 10
US-11-078-189-14
; Sequence 14, Application US/11078189
; Publication No. US20050277167A1
; GENERAL INFORMATION:
;  APPLICANT: Bachinger, Hans Peter
;  APPLICANT:  Vranka, Janice
;  TITLE OF INVENTION: PROLYL 3-HYDROXYLASES
;  FILE REFERENCE: 08062-020001
;  CURRENT APPLICATION NUMBER: US/11/078,189
;  CURRENT FILING DATE:  2005-03-11
;  PRIOR APPLICATION NUMBER: US 60/552,409
;  PRIOR FILING DATE: 2004-03-11
;  NUMBER OF SEQ ID NOS: 26
;  SOFTWARE: FastSEQ for Windows Version 4.0
; SEQ ID NO 14
;   LENGTH: 732
;   TYPE: PRT
;   ORGANISM: Mus musculus
US-11-078-189-14

  Query Match             1.4%;  Score 9;  DB 7;  Length 732;
  Best Local Similarity   100.0%;  Pred. No. 1.1;
  Matches    9;  Conservative    0;  Mismatches    0;  Indels    0;  Gaps    0;

Qy        383 EEEEEEEED 391
              |||||||||
Db        690 EEEEEEEED 698


RESULT 11
US-11-044-899-29
; Sequence 29, Application US/11044899
; Publication No. US20050260616A1
; GENERAL INFORMATION:
;  APPLICANT: Schwab, M.
;  APPLICANT:  Chen, M.
;  TITLE OF INVENTION: NUCLEOTIDE AND PROTEIN SEQUENCES OF NOGO GENES AND METHODS BASED
;  TITLE OF INVENTION:  THEREON
;  FILE REFERENCE: 10200-017-999
;  CURRENT APPLICATION NUMBER: US/11/044,899
;  CURRENT FILING DATE:  2005-01-26
;  PRIOR APPLICATION NUMBER: 09/830,972
;  PRIOR FILING DATE: 2001-09-24
;  PRIOR APPLICATION NUMBER: PCT/US99/26160
;  PRIOR FILING DATE: 1999-11-05
;  PRIOR APPLICATION NUMBER: 60/107,446
;  PRIOR FILING DATE: 1998-11-06
;  NUMBER OF SEQ ID NOS: 51
;  SOFTWARE: FastSEQ for Windows Version 3.0
; SEQ ID NO 29
;   LENGTH: 1178
;   TYPE: PRT
;   ORGANISM: Homo sapiens
;   FEATURE:
;   NAME/KEY: VARIANT
;   LOCATION: (1)...(1178) at all Xaa position
;   OTHER INFORMATION: Xaa = any amino acid
US-11-044-899-29

  Query Match             1.4%;  Score 9;  DB 7;  Length 1178;
  Best Local Similarity   100.0%;  Pred. No. 1.7;
  Matches    9;  Conservative    0;  Mismatches    0;  Indels    0;  Gaps    0;

Qy        383 EEEEEEEED 391
              |||||||||
Db         36 EEEEEEEED 44


RESULT 12
US-11-054-281-107
; Sequence 107, Application US/11054281
; Publication No. US20060013813A1
; GENERAL INFORMATION:
;  APPLICANT: Mezes et al.
;  TITLE OF INVENTION: Proteins and Nucleic Acids Encoding Same
;  FILE REFERENCE: 21402-240CIP
;  CURRENT APPLICATION NUMBER: US/11/054,281
;  CURRENT FILING DATE:  2005-02-08
;  PRIOR APPLICATION NUMBER: 60/261,014
;  PRIOR FILING DATE: 2001-01-11
;  PRIOR APPLICATION NUMBER: 60/261,018
;  PRIOR FILING DATE: 2001-01-11
;  PRIOR APPLICATION NUMBER: 60/318,410
;  PRIOR FILING DATE: 2001-09-10
;  PRIOR APPLICATION NUMBER: 60/261,013
;  PRIOR FILING DATE: 2001-01-11
;  PRIOR APPLICATION NUMBER: 60/261,026
;  PRIOR FILING DATE: 2001-01-11
;  PRIOR APPLICATION NUMBER: 60/261,029
;  PRIOR FILING DATE: 2001-01-11
;  PRIOR APPLICATION NUMBER: 60/313,170
;  PRIOR FILING DATE: 2001-08-17
;  PRIOR APPLICATION NUMBER: 10/044,564
;  PRIOR FILING DATE: 2002-01-11
;  NUMBER OF SEQ ID NOS: 324
;  SOFTWARE: PatentIn Ver. 2.1
; SEQ ID NO 107
;   LENGTH: 1229
;   TYPE: PRT
;   ORGANISM: Rattus norvegicus
US-11-054-281-107

  Query Match             1.4%;  Score 9;  DB 7;  Length 1229;
  Best Local Similarity   100.0%;  Pred. No. 1.7;
  Matches    9;  Conservative    0;  Mismatches    0;  Indels    0;  Gaps    0;

Qy        383 EEEEEEEED 391
              |||||||||
Db        152 EEEEEEEED 160


RESULT 13
US-11-126-313-31
; Sequence 31, Application US/11126313
; Publication No. US20050288489A1
; GENERAL INFORMATION:
;  APPLICANT: Hirsch, Joel
;  TITLE OF INVENTION: VOLTAGE-DEPENDENT CALCIUM CHANNEL BETA SUBUNIT FUNCTIONAL CORE
;  FILE REFERENCE: P-6758-US
;  CURRENT APPLICATION NUMBER: US/11/126,313
```

; CURRENT FILING DATE: 2005-05-11
; NUMBER OF SEQ ID NOS: 38
; SOFTWARE: PatentIn version 3.3
; SEQ ID NO 31
; LENGTH: 2161
; TYPE: PRT
; ORGANISM: homo sapiens
US-11-126-313-31

Query Match 1.4%; Score 9; DB 7; Length 2161;
Best Local Similarity 100.0%; Pred. No. 2.9;
Matches 9; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

Qy 383 EEEEEEEED 391
|||||||||
Db 827 EEEEEEEED 835

RESULT 14
US-11-044-111-24
; Sequence 24, Application US/11044111
; Publication No. US20050272362A1
; GENERAL INFORMATION:
; APPLICANT: Chiang, Wen
; APPLICANT: Strasburg, Gale
; APPLICANT: Linz, John
; TITLE OF INVENTION: Genetic Test for PSE-Susceptible Turkeys
; FILE REFERENCE: MSU-09308
; CURRENT APPLICATION NUMBER: US/11/044,111
; CURRENT FILING DATE: 2005-01-27
; NUMBER OF SEQ ID NOS: 27
; SOFTWARE: PatentIn version 3.3
; SEQ ID NO 24
; LENGTH: 4868
; TYPE: PRT
; ORGANISM: Meleagris gallopavo
US-11-044-111-24

Query Match 1.4%; Score 9; DB 7; Length 4868;
Best Local Similarity 100.0%; Pred. No. 6.3;
Matches 9; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

Qy 383 EEEEEEEED 391
|||||||||
Db 1915 EEEEEEEED 1923

RESULT 15
US-11-098-686-10232
; Sequence 10232, Application US/11098686
; Publication No. US20060024696A1
; GENERAL INFORMATION:
; APPLICANT: Kapur, Vivek and Gebhart, Connie J.
; TITLE OF INVENTION: NUCLEIC ACID AND POLYPEPTIDE SEQUENCES
; TITLE OF INVENTION: FROM LAWSONIA INTRACELLULARIS AND METHODS OF USING
; FILE REFERENCE: 09531-128001
; CURRENT APPLICATION NUMBER: US/11/098,686
; CURRENT FILING DATE: 2005-04-04
; PRIOR APPLICATION NUMBER: PCT/US03/31318
; PRIOR FILING DATE: 2003-10-01
; PRIOR APPLICATION NUMBER: US 60/416,395
; PRIOR FILING DATE: 2002-10-04
; NUMBER OF SEQ ID NOS: 11433
; SOFTWARE: FastSEQ for Windows Version 4.0
; SEQ ID NO 10232
; LENGTH: 8746
; TYPE: PRT
; ORGANISM: Lawsonia intracellularis
US-11-098-686-10232

Query Match 1.4%; Score 9; DB 7; Length 8746;
Best Local Similarity 100.0%; Pred. No. 11;
Matches 9; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

Qy 391 DYEDDKDDD 399
|||||||||
Db 2370 DYEDDKDDD 2378

Search completed: February 28, 2006, 09:27:39
Job time : 20 secs

GenCore version 5.1.7

OM protein - protein search, using sw model

Run on: February 28, 2006, 09:06:06 ; Search time 42 Seconds
(without alignments)
1463.869 Million cell updates/sec

Title: US-10-717-665A-44
Perfect score: 639
Sequence: 1 MPQPSVSGMDPPPFGDAFRSH.....TAEGNPTGGLVGLRIPTSKV 639

Scoring table: OLIGO
Gapop 60.0 , Gapext 60.0

Searched: 283416 seqs, 96216763 residues

Word size : 0

Total number of hits satisfying chosen parameters: 283416

Minimum DB seq length: 0
Maximum DB seq length: 2000000000

Post-processing: Listing first 45 summaries

Database : PIR_80:*
1: pir1:*
2: pir2:*
3: pir3:*
4: pir4:*

Pred. No. is the number of results predicted by chance to have a score greater than or equal to the score of the result being printed, and is derived by analysis of the total score distribution.

SUMMARIES

Result No.	Score	% Query Match	Length	DB	ID	Description
1	11	1.7	517	2	T23499	hypothetical prote
2	11	1.7	1479	2	T17401	transcription regu
3	10	1.6	208	2	S46510	hypothetical prote
4	10	1.6	547	2	T45635	hypothetical prote
5	10	1.6	571	2	S37913	hypothetical prote
6	10	1.6	786	2	A35466	progesterone recep
7	10	1.6	802	1	S48529	NAB3 protein - yea
8	10	1.6	1361	2	S50943	hypothetical prote
9	9	1.4	81	2	T48398	hypothetical prote
10	9	1.4	136	2	T30045	hypothetical prote
11	9	1.4	170	2	A27853	nonhistone chromos
12	9	1.4	184	2	D84746	hypothetical prote
13	9	1.4	186	2	S30221	nonhistone chromos
14	9	1.4	190	2	A26630	nucleoplasmin A -
15	9	1.4	200	2	A26169	nucleoplasmin - Af
16	9	1.4	202	2	A26301	dopamine- and cAMP
17	9	1.4	207	2	JC1129	nonhistone chromos
18	9	1.4	207	2	JC1114	high-mobility grou
19	9	1.4	209	1	NSHUH2	nonhistone chromos
20	9	1.4	210	2	A34719	nonhistone chromos
21	9	1.4	210	2	S54774	high mobility grou
22	9	1.4	215	1	NSRTH1	nonhistone chromos
23	9	1.4	215	1	S01947	nonhistone chromos
24	9	1.4	215	2	S02826	nonhistone chromos
25	9	1.4	215	2	A28897	nonhistone chromos
26	9	1.4	215	2	I48688	non-histone chromo
27	9	1.4	215	2	T05158	hypothetical prote
28	9	1.4	216	2	S29857	nonhistone chromos
29	9	1.4	221	2	A84638	hypothetical prote
30	9	1.4	254	2	T01109	hypothetical prote
31	9	1.4	267	2	H84857	hypothetical prote
32	9	1.4	271	2	G85035	hypothetical prote
33	9	1.4	294	2	E84706	hypothetical prote
34	9	1.4	295	2	T39012	hypothetical prote
35	9	1.4	302	2	E86267	hypothetical prote
36	9	1.4	336	2	G69091	ribosomal protein
37	9	1.4	343	2	T02399	hypothetical prote
38	9	1.4	350	2	S00337	legumin B LegK pre
39	9	1.4	357	2	I49338	neurogenic differe
40	9	1.4	357	2	JC4703	basic helix-loop-h
41	9	1.4	366	2	A96692	hypothetical prote
42	9	1.4	368	2	H96712	probable DNA-bindi
43	9	1.4	372	2	T04266	hypothetical prote
44	9	1.4	376	2	T10455	heat shock related
45	9	1.4	381	2	T49544	hypothetical prote

ALIGNMENTS

RESULT 1
T23499
hypothetical protein K08F9.4 - Caenorhabditis elegans
C;Species: Caenorhabditis elegans
C;Date: 15-Oct-1999 #sequence_revision 15-Oct-1999 #text_change 09-Jul-2004
C;Accession: T23499; T24766
R;Mortimore, B.
submitted to the EMBL Data Library, October 1996
A;Reference number: Z19748
A;Accession: T23499
A;Status: preliminary; translated from GB/EMBL/DDBJ
A;Molecule type: DNA
A;Residues: 1-517 <WIL>
A;Cross-references: UNIPROT:O17941; UNIPARC:UPI000007CD27; EMBL:Z81099; PIDN:CAB03189.1;
A;Experimental source: clone K08F9
R;Mortimore, B.
submitted to the EMBL Data Library, December 1996
A;Reference number: Z19933
A;Accession: T24766
A;Status: preliminary; translated from GB/EMBL/DDBJ
A;Molecule type: DNA
A;Residues: 1-517 <WI2>
A;Cross-references: UNIPARC:UPI000007CD27; EMBL:Z83239; PIDN:CAB05811.1; GSPDB:GN00023; (
A;Experimental source: clone T09F5
C;Genetics:
A;Gene: CESP:K08F9.4
A;Map position: 5
A;Introns: 16/2; 44/2; 66/2; 277/2; 341/3; 410/2; 426/3
C;Superfamily: Caenorhabditis elegans hypothetical protein K08F9.4

Query Match 1.7%; Score 11; DB 2; Length 517;
Best Local Similarity 100.0%; Pred. No. 0.017;
Matches 11; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

```
Qy        380 SENEEEEEEEE 390
              ||||||||||
Db         21 SENEEEEEEEE 31
```

RESULT 2
T17401
transcription regulator WBSCR9 - mouse
N;Alternate names: Williams-Beuren syndrome deletion transcript 9 homolog
C;Species: Mus musculus (house mouse)
C;Date: 15-Oct-1999 #sequence_revision 15-Oct-1999 #text_change 09-Jul-2004
C;Accession: T17401
R;Peoples, R.J.; Cisco, M.J.; Kaplan, P.; Francke, U.
Cytogenet. Cell Genet. 82, 238-246, 1998
A;Title: Identification of the WBSCR9 gene, encoding a novel transcriptional regulator, i
A;Reference number: Z18735; MUID:99077764; PMID:9858827
A;Accession: T17401
A;Status: preliminary; translated from GB/EMBL/DDBJ

A;Molecule type: mRNA
A;Residues: 1-1479 <PEO>
A;Cross-references: UNIPROT:Q9Z277; UNIPARC:UPI0000029780; EMBL:AF084480; NID:g4165088;
C;Genetics:
A;Gene: Wbscr9
A;Map position: 5
F;1360-1415/Domain: bromodomain homology <BRO>

Query Match 1.7%; Score 11; DB 2; Length 1479;
Best Local Similarity 100.0%; Pred. No. 0.044;
Matches 11; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

Qy 383 EEEEEEEEDYE 393
||||||||||
Db 1267 EEEEEEEEDYE 1277

RESULT 3
S46510
hypothetical protein - Agrobacterium tumefaciens
C;Species: Agrobacterium tumefaciens
C;Date: 26-Dec-1994 #sequence_revision 26-May-1995 #text_change 09-Jul-2004
C;Accession: S46510
R;Drevet, C.; Brasileiro, A.C.M.; Jouanin, L.
Plant Mol. Biol. 25, 83-90, 1994
A;Title: Oncogene arrangement in a shooty strain of Agrobacterium tumefaciens.
A;Reference number: S46509; MUID:94272016; PMID:8003699
A;Accession: S46510
A;Status: preliminary; nucleic acid sequence not shown; translation not shown
A;Molecule type: DNA
A;Residues: 1-208 <DRE>
A;Cross-references: UNIPROT:Q57530; UNIPARC:UPI00000BEEC9; EMBL:X74123; NID:g510732; PID
A;Note: the nucleotide sequence was submitted to the EMBL Data Library, July 1993
C;Superfamily: T-6b protein

Query Match 1.6%; Score 10; DB 2; Length 208;
Best Local Similarity 100.0%; Pred. No. 0.078;
Matches 10; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

Qy 381 ENEEEEEEEE 390
||||||||||
Db 173 ENEEEEEEEE 182

RESULT 4
T45635
hypothetical protein F13I12.4 - Arabidopsis thaliana
N;Alternate names: protein F13I12.40
C;Species: Arabidopsis thaliana (mouse-ear cress)
C;Date: 04-Feb-2000 #sequence_revision 04-Feb-2000 #text_change 09-Jul-2004
C;Accession: T45635
R;Choisne, N.; Robert, C.; Brottier, P.; Wincker, P.; Cattolico, L.; Artiguenave, F.; Sa
submitted to the Protein Sequence Database, November 1999
A;Reference number: Z23010
A;Accession: T45635
A;Status: preliminary
A;Molecule type: DNA
A;Residues: 1-547 <CHO>
A;Cross-references: UNIPROT:Q9SD74; UNIPARC:UPI00000A6868; EMBL:AL133292
A;Experimental source: cultivar Columbia; BAC clone F13I12
C;Genetics:
A;Gene: F13I12.40
A;Map position: 3

Query Match 1.6%; Score 10; DB 2; Length 547;
Best Local Similarity 100.0%; Pred. No. 0.19;
Matches 10; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

Qy 383 EEEEEEEEDY 392
||||||||||
Db 116 EEEEEEEEDY 125

RESULT 5
S37913
hypothetical protein YKL088w - yeast (Saccharomyces cerevisiae)
C;Species: Saccharomyces cerevisiae
C;Date: 03-May-1994 #sequence_revision 03-May-1994 #text_change 09-Jul-2004
C;Accession: S37913
R;Pohl, T.M.; Pohl, F.M.
submitted to the Protein Sequence Database, March 1994
A;Reference number: S37897
A;Accession: S37913
A;Molecule type: DNA
A;Residues: 1-571 <POH>
A;Cross-references: UNIPROT:P36076; UNIPARC:UPI0000052F06; EMBL:Z28088; NID:g486130; PID:
A;Experimental source: strain S288C
C;Genetics:
A;Gene: MIPS:YKL088w
A;Cross-references: SGD:S0001571
A;Map position: 11L

Query Match 1.6%; Score 10; DB 2; Length 571;
Best Local Similarity 100.0%; Pred. No. 0.2;
Matches 10; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

Qy 355 EEDEEDEEDV 364
||||||||||
Db 558 EEDEEDEEDV 567

RESULT 6
A35466
progesterone receptor form B - chicken
N;Contains: progesterone receptor form A
C;Species: Gallus gallus (chicken)
C;Date: 18-Nov-1994 #sequence_revision 18-Nov-1994 #text_change 09-Jul-2004
C;Accession: A35466; S06284; A40903; A24661; A24312; A40911; A61552
R;Jeltsch, J.M.; Turcotte, B.; Garnier, J.M.; Lerouge, T.; Krozowski, Z.; Gronemeyer, H.;
J. Biol. Chem. 265, 3967-3974, 1990
A;Title: Characterization of multiple mRNAs originating from the chicken progesterone rec
A;Reference number: A35466; MUID:90154085; PMID:2303488
A;Accession: A35466
A;Status: translation not shown
A;Molecule type: DNA
A;Residues: 1-786 <JEL>
A;Cross-references: UNIPROT:P07812; UNIPARC:UPI00001321A4; GB:M32732; GB:J05240; NID:g21
R;Gronemeyer, H.; Turcotte, B.; Quirin-Stricker, C.; Bocquel, M.T.; Meyer, M.E.; Krozowsk
EMBO J. 6, 3985-3994, 1987
A;Title: The chicken progesterone receptor: sequence, expression and functional analysis.
A;Reference number: S06284; MUID:88166640; PMID:3443098
A;Accession: S06284
A;Molecule type: DNA
A;Residues: 1-786 <GRO>
A;Cross-references: UNIPARC:UPI00001321A4; EMBL:Y00092; NID:g63744; PIDN:CAA68282.1; PID:
R;Conneely, O.M.; Dobson, A.D.W.; Tsai, M.J.; Beattie, W.G.; Toft, D.O.; Huckaby, C.S.; Z
Mol. Endocrinol. 1, 517-525, 1987
A;Title: Sequence and expression of a functional chicken progesterone receptor.
A;Reference number: A40903; MUID:91042592; PMID:3153474
A;Accession: A40903
A;Status: preliminary
A;Molecule type: mRNA
A;Residues: 1-64,'E',65-786 <CON>
A;Cross-references: UNIPARC:UPI0000177C09; GB:M37518
R;Conneely, O.M.; Sullivan, W.P.; Toft, D.O.; Birnbaumer, M.; Cook, R.G.; Maxwell, B.L.;
Science 233, 767-770, 1986
A;Title: Molecular cloning of the chicken progesterone receptor.
A;Reference number: A24661; MUID:86289413; PMID:2426779
A;Accession: A24661
A;Molecule type: mRNA
A;Residues: 128-133,'E',135-147,'E',149-164 <CO2>
A;Cross-references: UNIPARC:UPI0000177C0A
A;Note: the authors translated the codon CAG for residue 7 as Glu and CAG for residue 21
R;Jeltsch, J.M.; Krozowski, Z.; Quirin-Stricker, C.; Gronemeyer, H.; Simpson, R.J.; Garni
Proc. Natl. Acad. Sci. U.S.A. 83, 5424-5428, 1986
A;Title: Cloning of the chicken progesterone receptor.

```
A;Reference number: A24312; MUID:86287271; PMID:2426697
A;Accession: A24312
A;Molecule type: mRNA
A;Residues: 417-490 <JE2>
A;Cross-references: UNIPARC:UPI000017139B; GB:M14280; NID:g212607; PIDN:AAA49039.1; PID:
A;Note: amino acid and corresponding nucleotide sequences are also shown for three small
R;Birnbaumer, M.; Hinrichs-Rosello, M.V.; Cook, R.G.; Schrader, W.T.; O'Malley, B.W.
Mol. Endocrinol. 1, 249-259, 1987
A;Title: Chemical and antigenic properties of pure 108,000 molecular weight chick proges
A;Reference number: A40911; MUID:88288199; PMID:3453892
A;Accession: A40911
A;Status: preliminary
A;Molecule type: protein
A;Residues: 128-133,'E',135-147,'E',149-164;546-558 <BIR>
A;Cross-references: UNIPARC:UPI0000177C0A; UNIPARC:UPI0000177C0B
R;Simpson, R.J.; Grego, B.; Govindan, M.V.; Gronemeyer, H.
Mol. Cell. Endocrinol. 52, 177-184, 1987
A;Title: Peptide sequencing of the chick oviduct progesterone receptor form B.
A;Reference number: A61552; MUID:88005426; PMID:3653503
A;Accession: A61552
A;Molecule type: protein
A;Residues: 136-153;168-174;195-228;526-537,'X',539;546-563 <SIM>
A;Cross-references: UNIPARC:UPI0000177C0C; UNIPARC:UPI0000177C0D; UNIPARC:UPI0000177C0E;
C;Genetics:
A;Introns: 400/2; 451/1; 490/1; 591/1; 639/2; 683/1; 735/3
C;Superfamily: progesterone receptor; erbA transforming protein homology
C;Keywords: DNA binding; nucleus; steroid hormone receptor; zinc finger
F;1-786/Product: progesterone receptor form B #status predicted <MA1>
F;128-786/Product: progesterone receptor form A #status predicted <MA2>
F;419-682/Domain: erbA transforming protein homology <ERBA>
F;421-441/Region: zinc finger
F;457-481/Region: zinc finger

  Query Match             1.6%;  Score 10;  DB 2;  Length 786;
  Best Local Similarity   100.0%;  Pred. No. 0.26;
  Matches   10;  Conservative    0;  Mismatches    0;  Indels    0;  Gaps    0;

Qy        381 ENEEEEEEEE 390
              ||||||||||
Db         55 ENEEEEEEEE 64


RESULT 7
S48529
NAB3 protein - yeast (Saccharomyces cerevisiae)
N;Alternate names: probable RNA/ssDNA-binding protein HMD1; protein P1945; protein YPL19
C;Species: Saccharomyces cerevisiae
C;Date: 10-Sep-1999 #sequence_revision 10-Sep-1999 #text_change 09-Jul-2004
C;Accession: S48529; S65209; S60122
R;Wilson, S.M.; Oberdorf, A.M.; Datar, K.V.; Swedlow, J.R.; Paddy, M.R.; Swanson, M.S.
submitted to the EMBL Data Library, January 1994
A;Description: Characterization of Nuclear Polyadenylated RNA-Binding Proteins from Sacc
A;Reference number: S48529
A;Accession: S48529
A;Molecule type: DNA
A;Residues: 1-802 <WIL>
A;Cross-references: UNIPROT:P38996; UNIPARC:UPI000004F979; EMBL:U05314; NID:g476219; PID
R;Rieger, M.; Mueller-Auer, S.; Schaefer, M.
submitted to the Protein Sequence Database, May 1996
A;Reference number: S65202
A;Accession: S65209
A;Molecule type: DNA
A;Residues: 1-802 <RIE>
A;Cross-references: UNIPARC:UPI000004F979; EMBL:Z73546; NID:g1370396; PID:g1370397; GSPD
A;Experimental source: strain S288C (AB972)
R;Sugimoto, K.; Matsumoto, K.; Kornberg, R.D.; Reed, S.I.; Wittenberg, C.
Mol. Gen. Genet. 248, 712-718, 1995
A;Title: Dosage suppressors of the dominant G1 cyclin mutant CLN3-2: identification of a
A;Reference number: S60122; MUID:96069710; PMID:7476874
A;Accession: S60122
A;Status: nucleic acid sequence not shown
A;Molecule type: DNA
A;Residues: 1-340,'I',342-802 <SUG>
A;Cross-references: UNIPARC:UPI00000694F8; GB:D37935; NID:g1235749; PID:g1235750
C;Genetics:
A;Gene: SGD:NAB3; NAB3; HMD1; MIPS:YPL190c
A;Cross-references: SGD:S0006111
A;Map position: 16L
C;Superfamily: yeast NAB3 protein; ribonucleoprotein repeat homology
C;Keywords: nucleus; RNA binding
F;331-396/Domain: ribonucleoprotein repeat homology <RRM1>

  Query Match             1.6%;  Score 10;  DB 1;  Length 802;
  Best Local Similarity   100.0%;  Pred. No. 0.27;
  Matches   10;  Conservative    0;  Mismatches    0;  Indels    0;  Gaps    0;

Qy        382 NEEEEEEEED 391
              ||||||||||
Db        107 NEEEEEEEED 116


RESULT 8
S50943
hypothetical protein YML049c - yeast (Saccharomyces cerevisiae)
N;Alternate names: hypothetical protein YM9827.03c
C;Species: Saccharomyces cerevisiae
C;Date: 10-Feb-1995 #sequence_revision 12-May-1995 #text_change 09-Jul-2004
C;Accession: S50943
R;Odell, C.; Bowman, S.
submitted to the EMBL Data Library, January 1995
A;Reference number: S50941
A;Accession: S50943
A;Molecule type: DNA
A;Residues: 1-1361 <ODE>
A;Cross-references: UNIPROT:Q04693; UNIPARC:UPI00000530F0; EMBL:Z47816; NID:g642303; PIDN
C;Genetics:
A;Gene: SGD:RSE1
A;Cross-references: SGD:S0004513; MIPS:YML049c
A;Map position: 13L

  Query Match             1.6%;  Score 10;  DB 2;  Length 1361;
  Best Local Similarity   100.0%;  Pred. No. 0.43;
  Matches   10;  Conservative    0;  Mismatches    0;  Indels    0;  Gaps    0;

Qy        382 NEEEEEEEED 391
              ||||||||||
Db        789 NEEEEEEEED 798


RESULT 9
T48398
hypothetical protein F17C15.130 - Arabidopsis thaliana
C;Species: Arabidopsis thaliana (mouse-ear cress)
C;Date: 20-Apr-2000 #sequence_revision 20-Apr-2000 #text_change 09-Jul-2004
C;Accession: T48398
R;Bevan, M.; Pohl, T.; Weizenegger, T.; Bancroft, I.; Mewes, H.W.; Rudd, S.; Lemcke, K.;
submitted to the Protein Sequence Database, March 2000
A;Reference number: Z24492
A;Accession: T48398
A;Status: preliminary
A;Molecule type: DNA
A;Residues: 1-81 <BEV>
A;Cross-references: UNIPROT:Q9LZR7; UNIPARC:UPI00000AAEF5; EMBL:AL162506
A;Experimental source: cultivar Columbia; BAC clone F17C15
C;Genetics:
A;Map position: 5
A;Introns: 5/1
A;Note: F17C15.130

  Query Match             1.4%;  Score 9;  DB 2;  Length 81;
  Best Local Similarity   100.0%;  Pred. No. 0.35;
  Matches    9;  Conservative    0;  Mismatches    0;  Indels    0;  Gaps    0;

Qy        383 EEEEEEEED 391
              |||||||||
Db         59 EEEEEEEED 67
```

```
RESULT 10
T30045
hypothetical protein C16H3.4 - Caenorhabditis elegans
C;Species: Caenorhabditis elegans
C;Date: 15-Oct-1999 #sequence_revision 15-Oct-1999 #text_change 09-Jul-2004
C;Accession: T30045
R;Geisel, C.; Bradshaw, H.
submitted to the EMBL Data Library, August 1996
A;Description: The sequence of C. elegans cosmid C16H3.
A;Reference number: Z20727
A;Accession: T30045
A;Status: preliminary; translated from GB/EMBL/DDBJ
A;Molecule type: DNA
A;Residues: 1-136 <GEI>
A;Cross-references: UNIPROT:Q94170; UNIPARC:UPI000017B78F; EMBL:U67955; PIDN:AAB07583.1;
A;Experimental source: strain Bristol N2; clone C16H3
C;Genetics:
A;Gene: CESP:C16H3.4
A;Map position: X
A;Introns: 26/1

  Query Match             1.4%;  Score 9;  DB 2;  Length 136;
  Best Local Similarity   100.0%;  Pred. No. 0.56;
  Matches    9;  Conservative    0;  Mismatches    0;  Indels    0;  Gaps    0;

Qy        383 EEEEEEEED 391
              |||||||||
Db         98 EEEEEEEED 106


RESULT 11
A27853
nonhistone chromosomal protein HMG-1 - Chinese hamster (fragment)
C;Species: Cricetulus griseus (Chinese hamster)
C;Date: 30-Jun-1989 #sequence_revision 30-Jun-1989 #text_change 09-Jul-2004
C;Accession: A27853
R;Lee, K.L.D.; Pentecost, B.T.; D'Anna, J.A.; Tobey, R.A.; Gurley, L.R.; Dixon, G.H.
Nucleic Acids Res. 15, 5051-5068, 1987
A;Title: Characterization of cDNA sequences corresponding to three distinct HMG-1 mRNA s
A;Reference number: A27853; MUID:87259986; PMID:3601666
A;Accession: A27853
A;Molecule type: mRNA
A;Residues: 1-170 <LEE>
A;Cross-references: UNIPROT:P07156; UNIPARC:UPI00001771D5
C;Superfamily: nonhistone chromosomal protein HMG-2; HMG box homology
C;Keywords: chromosomal protein; DNA binding; nucleus
F;1-38/Domain: HMG box homology (fragment) <HMG1>
F;47-121/Domain: HMG box homology <HMG2>

  Query Match             1.4%;  Score 9;  DB 2;  Length 170;
  Best Local Similarity   100.0%;  Pred. No. 0.69;
  Matches    9;  Conservative    0;  Mismatches    0;  Indels    0;  Gaps    0;

Qy        383 EEEEEEEED 391
              |||||||||
Db        152 EEEEEEEED 160


RESULT 12
D84746
hypothetical protein At2g33510 [imported] - Arabidopsis thaliana
C;Species: Arabidopsis thaliana (mouse-ear cress)
C;Date: 02-Feb-2001 #sequence_revision 02-Feb-2001 #text_change 09-Jul-2004
C;Accession: D84746
R;Lin, X.; Kaul, S.; Rounsley, S.D.; Shea, T.P.; Benito, M.I.; Town, C.D.; Fujii, C.Y.;
M.; Koo, H.; Moffat, K.S.; Cronin, L.A.; Shen, M.; VanAken, S.E.; Umayam, L.; Tallon, L.
euss, D.; Nierman, W.C.; White, O.; Eisen, J.A.; Salzberg, S.L.; Fraser, C.M.; Venter, J
Nature 402, 761-768, 1999
A;Title: Sequence and analysis of chromosome 2 of the plant Arabidopsis thaliana.
A;Reference number: A84420; MUID:20083487; PMID:10617197
A;Accession: D84746
A;Status: preliminary
A;Molecule type: DNA
A;Residues: 1-184 <STO>
A;Cross-references: UNIPROT:O22801; UNIPARC:UPI00000A10DC; GB:AE002093; NID:g2459433; PII
C;Genetics:
A;Gene: At2g33510
A;Map position: 2

  Query Match             1.4%;  Score 9;  DB 2;  Length 184;
  Best Local Similarity   100.0%;  Pred. No. 0.74;
  Matches    9;  Conservative    0;  Mismatches    0;  Indels    0;  Gaps    0;

Qy        383 EEEEEEEED 391
              |||||||||
Db        135 EEEEEEEED 143


RESULT 13
S30221
nonhistone chromosomal protein HMG-2B - human (fragment)
C;Species: Homo sapiens (man)
C;Date: 22-Nov-1993 #sequence_revision 26-May-1995 #text_change 09-Jul-2004
C;Accession: S30221
R;Alexandre, S.; Li, W.W.; Lee, A.S.
Nucleic Acids Res. 20, 6413, 1992
A;Title: A human HMG2 cDNA with a novel 3'-untranslated region.
A;Reference number: S30221; MUID:93117123; PMID:1475204
A;Accession: S30221
A;Status: preliminary; nucleic acid sequence not shown; translation not shown
A;Molecule type: mRNA
A;Residues: 1-186 <ALE>
A;Cross-references: UNIPROT:P26583; UNIPARC:UPI000016AA6C; EMBL:Z17240; NID:g32334; PIDN:
A;Note: the nucleotide sequence was submitted to the EMBL Data Library, October 1992
C;Superfamily: nonhistone chromosomal protein HMG-2; HMG box homology
C;Keywords: chromosomal protein
F;1-60/Domain: HMG box homology (fragment) <HMG>
F;69-143/Domain: HMG box homology <HMG1>

  Query Match             1.4%;  Score 9;  DB 2;  Length 186;
  Best Local Similarity   100.0%;  Pred. No. 0.75;
  Matches    9;  Conservative    0;  Mismatches    0;  Indels    0;  Gaps    0;

Qy        383 EEEEEEEED 391
              |||||||||
Db        167 EEEEEEEED 175


RESULT 14
A26630
nucleoplasmin A - African clawed frog (fragment)
C;Species: Xenopus laevis (African clawed frog)
C;Date: 19-Nov-1988 #sequence_revision 19-Nov-1988 #text_change 09-Jul-2004
C;Accession: A26630
R;Burglin, T.R.; Mattaj, I.W.; Newmeyer, D.D.; Zeller, R.; De Robertis, E.M.
Genes Dev. 1, 97-107, 1987
A;Title: Cloning of nucleoplasmin from Xenopus laevis oocytes and analysis of its develop
A;Reference number: A26630; MUID:88112783; PMID:3428591
A;Accession: A26630
A;Molecule type: mRNA
A;Residues: 1-190 <BUR>
A;Cross-references: UNIPROT:P05221; UNIPARC:UPI0000171574; GB:Y00204; NID:g64938; PIDN:C
C;Superfamily: nucleophosmin
C;Keywords: molecular chaperone; nucleus

  Query Match             1.4%;  Score 9;  DB 2;  Length 190;
  Best Local Similarity   100.0%;  Pred. No. 0.76;
  Matches    9;  Conservative    0;  Mismatches    0;  Indels    0;  Gaps    0;

Qy        383 EEEEEEEED 391
              |||||||||
Db        128 EEEEEEEED 136
```

RESULT 15
A26169
nucleoplasmin - African clawed frog
C;Species: Xenopus laevis (African clawed frog)
C;Date: 19-Nov-1988 #sequence_revision 19-Nov-1988 #text_change 09-Jul-2004
C;Accession: A26169
R;Dingwall, C.; Dilworth, S.M.; Black, S.J.; Kearsey, S.E.; Cox, L.S.; Laskey, R.A.
EMBO J. 6, 69-74, 1987
A;Title: Nucleoplasmin cDNA sequence reveals polyglutamic acid tracts and a cluster of s
A;Reference number: A26169; MUID:87218476; PMID:2884102
A;Accession: A26169
A;Molecule type: mRNA
A;Residues: 1-200 <DIN>
A;Cross-references: UNIPROT:P05221; UNIPARC:UPI0000130A30; GB:X04766; NID:g64939; PIDN:C
C;Superfamily: nucleophosmin
C;Keywords: molecular chaperone; nucleus

Query Match 1.4%; Score 9; DB 2; Length 200;
Best Local Similarity 100.0%; Pred. No. 0.8;
Matches 9; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

```
Qy        383 EEEEEEEED 391
              |||||||||
Db        138 EEEEEEEED 146
```

Search completed: February 28, 2006, 09:10:32
Job time : 43 secs

GenCore version 5.1.7

OM protein - protein search, using sw model

Run on: February 28, 2006, 09:02:56 ; Search time 230 Seconds
(without alignments)
1960.142 Million cell updates/sec

Title: US-10-717-665A-44
Perfect score: 639
Sequence: 1 MPQPSVSGMDPPFGDAFRSH.....TAEGNPTGGLVGLRIPTSKV 639

Scoring table: OLIGO
Gapop 60.0 , Gapext 60.0

Searched: 2166443 seqs, 705528306 residues

Word size : 0

Total number of hits satisfying chosen parameters: 2166443

Minimum DB seq length: 0
Maximum DB seq length: 2000000000

Post-processing: Listing first 45 summaries

Database : UniProt_05.80:*
1: uniprot_sprot:*
2: uniprot_trembl:*

Pred. No. is the number of results predicted by chance to have a score greater than or equal to the score of the result being printed, and is derived by analysis of the total score distribution.

SUMMARIES

Result No.	Score	% Query Match	Length	DB	ID	Description
1	433	67.8	639	2	Q8IZG1_HUMAN	Q8izg1 homo sapien
2	381	59.6	639	2	Q8IUR6_HUMAN	Q8iur6 homo sapien
3	307	48.0	408	2	Q5HYG4_HUMAN	Q5hyg4 homo sapien
4	307	48.0	417	2	Q86YR3_HUMAN	Q86yr3 homo sapien
5	272	42.6	604	2	Q5HYK0_HUMAN	Q5hyk0 homo sapien
6	87	13.6	330	2	Q9CTQ7_MOUSE	Q9ctq7 mus musculu
7	87	13.6	351	2	Q8C8N7_MOUSE	Q8c8n7 mus musculu
8	87	13.6	640	2	Q8CDG5_MOUSE	Q8cdg5 mus musculu
9	44	6.9	131	2	Q5ZMD5_CHICK	Q5zmd5 gallus gall
10	34	5.3	600	2	Q4RRX3_TETNG	Q4rrx3 tetraodon n
11	11	1.7	167	2	Q6UAY4_WSSV	Q6uay4 white spot
12	11	1.7	215	2	Q9QTF1_9BACU	Q9qtf1 unidentifie
13	11	1.7	358	2	Q91B56_WSSV	Q91b56 white spot
14	11	1.7	468	2	Q6BKF1_DEBHA	Q6bkf1 debaryomyce
15	11	1.7	517	2	O17941_CAEEL	O17941 caenorhabdi
16	11	1.7	657	2	Q8CAU9_MOUSE	Q8cau9 mus musculu
17	11	1.7	1100	2	Q91LA3_WSSV	Q91la3 white spot
18	11	1.7	1179	2	Q91L98_WSSV	Q91l98 white spot
19	11	1.7	1180	2	Q8VAS9_WSSV	Q8vas9 white spot
20	11	1.7	1185	2	Q8QTC5_WSSV	Q8qtc5 white spot
21	11	1.7	1479	1	BAZ1B_MOUSE	Q9z277 mus musculu
22	11	1.7	1483	1	BAZ1B_HUMAN	Q9uig0 homo sapien
23	10	1.6	108	2	Q9R465_9RHIZ	Q9r465 agrobacteri
24	10	1.6	162	2	Q9LTZ3_ARATH	Q9ltz3 arabidopsis
25	10	1.6	196	2	Q54IM1_DICDI	Q54im1 dictyosteli
26	10	1.6	207	2	Q6AVQ0_ORYSA	Q6avq0 oryza sativ
27	10	1.6	208	2	Q57530_9RHIZ	Q57530 agrobacteri
28	10	1.6	219	2	Q4XM08_PLACH	Q4xm08 plasmodium
29	10	1.6	236	2	Q96IS5_HUMAN	Q96is5 homo sapien
30	10	1.6	255	2	Q6UW11_HUMAN	Q6uw11 homo sapien
31	10	1.6	273	2	Q96D96_HUMAN	Q96d96 homo sapien
32	10	1.6	289	2	Q5A212_CANAL	Q5a212 candida alb
33	10	1.6	293	2	Q7QG21_ANOGA	Q7qg21 anopheles g
34	10	1.6	297	2	Q707X8_KLUMA	Q707x8 kluyveromyc
35	10	1.6	340	2	Q9U7C9_DICDI	Q9u7c9 dictyosteli
36	10	1.6	361	2	Q54J82_DICDI	Q54j82 dictyosteli
37	10	1.6	432	2	Q5B7F4_EMENI	Q5b7f4 aspergillus
38	10	1.6	473	2	Q6BGT2_DEBHA	Q6bgt2 debaryomyce
39	10	1.6	547	2	Q9SD74_ARATH	Q9sd74 arabidopsis
40	10	1.6	571	1	YKI8_YEAST	P36076 saccharomyc
41	10	1.6	571	2	Q54RS0_DICDI	Q54rs0 dictyosteli
42	10	1.6	589	2	Q5I858_HUMAN	Q5i858 homo sapien
43	10	1.6	593	1	ARI3A_HUMAN	Q99856 homo sapien
44	10	1.6	593	2	Q6P9C6_HUMAN	Q6p9c6 homo sapien
45	10	1.6	731	2	Q5A868_CANAL	Q5a868 candida alb

ALIGNMENTS

RESULT 1
Q8IZG1_HUMAN
ID Q8IZG1_HUMAN PRELIMINARY; PRT; 639 AA.
AC Q8IZG1;
DT 01-MAR-2003 (TrEMBLrel. 23, Created)
DT 01-MAR-2003 (TrEMBLrel. 23, Last sequence update)
DT 01-MAR-2004 (TrEMBLrel. 26, Last annotation update)
DE Adult retina protein.
OS Homo sapiens (Human).
OC Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;
OC Mammalia; Eutheria; Euarchontoglires; Primates; Catarrhini; Hominidae;
OC Homo.
OX NCBI_TaxID=9606;
RN [1]
RP NUCLEOTIDE SEQUENCE.
RA Li F., Yao K.T.;
RL Submitted (AUG-2002) to the EMBL/GenBank/DDBJ databases.
DR EMBL; AY139008; AAN28956.1; -; mRNA.
DR InterPro; IPR004827; TF_bZIP.
DR PROSITE; PS00036; BZIP_BASIC; UNKNOWN_1.
SQ SEQUENCE 639 AA; 72090 MW; 3A19E0926B9A6406 CRC64;

Query Match 67.8%; Score 433; DB 2; Length 639;
Best Local Similarity 99.8%; Pred. No. 0;
Matches 533; Conservative 0; Mismatches 1; Indels 0; Gaps 0;

```
Qy          1 MPQPSVSGMDPPFGDAFRSHTFSEQTLMSTDLLANSSDPDFMYELDREMNYQQNPRDNFL 60
              ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db          1 MPQPSVSGMDPPFGDAFRSHTFSEQTLMSTDLLANSSDPDFMYELDREMNYQQNPRDNFL 60

Qy         61 SLEDCKDIENLESFTDVLDNEGALTSNWEQWDTYCEDLTKYTKLTSCDIWGTKEVDYLGL 120
              ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db         61 SLEDCKDIENLESFTDVLDNEGALTSNWEQWDTYCEDLTKYTKLTSCDIWGTKEVDYLGL 120

Qy        121 DDFSSPYQDEEVISKTPTLAQLNSEDSQSVSDSLYYPDSLFSVKQNPLPSSFPGKKITSR 180
              ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db        121 DDFSSPYQDEEVISKTPTLAQLNSEDSQSVSDSLYYPDSLFSVKQNPLPSSFPGKKITSR 180

Qy        181 AAAPVCSSKTLQAEVPLSDCVQKASKPPSSTQIMVKTNMYHNEKVNFHVECKDYVKKAKV 240
              ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db        181 AAAPVCSSKTLQAEVPLSDCVQKASKPTSSTQIMVKTNMYHNEKVNFHVECKDYVKKAKV 240

Qy        241 KINPVQQSRPLLSQIHTDAAKENTCYCGAVAKRQEKKGMEPLQGHATPALPFKETQELLL 300
              ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db        241 KINPVQQSRPLLSQIHTDAAKENTCYCGAVAKRQEKKGMEPLQGHATPALPFKETQELLL 300

Qy        301 SPLPQEGPGSLAAGESSSLSASTSVSDSSQKKEEHNYSLFVSDNLGEQPTKCSPEEDEED 360
              ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db        301 SPLPQEGPGSLAAGESSSLSASTSVSDSSQKKEEHNYSLFVSDNLGEQPTKCSPEEDEED 360

Qy        361 EEDVDDEDHDEGFGSEHELSENEEEEEEEEDYEDDKDDDISDTFSEPGYENDSVEDLKEV 420
              ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db        361 EEDVDDEDHDEGFGSEHELSENEEEEEEEEDYEDDKDDDISDTFSEPGYENDSVEDLKEV 420
```

```
Qy        421 TSISSRKRGKRRYFWEYSEQLTPSQQERMLRPSEWNRDTLPSNMYQKNGLHHGKYAVKKS 480
              ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db        421 TSISSRKRGKRRYFWEYSEQLTPSQQERMLRPSEWNRDTLPSNMYQKNGLHHGKYAVKKS 480

Qy        481 RRTDVEDLTPNPKKLLQIGNELRKLNKVISDLTPVSELPLTARPRSRKEKNKLA 534
              |||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db        481 RRTDVEDLTPNPKKLLQIGNELRKLNKVISDLTPVSELPLTARPRSRKEKNKLA 534


RESULT 2
Q8IUR6_HUMAN
ID   Q8IUR6_HUMAN PRELIMINARY;      PRT;   639 AA.
AC   Q8IUR6;
DT   01-MAR-2003 (TrEMBLrel. 23, Created)
DT   01-MAR-2003 (TrEMBLrel. 23, Last sequence update)
DT   01-MAR-2004 (TrEMBLrel. 26, Last annotation update)
DE   Adult retina protein.
GN   Name=LOC153222;
OS   Homo sapiens (Human).
OC   Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;
OC   Mammalia; Eutheria; Euarchontoglires; Primates; Catarrhini; Hominidae;
OC   Homo.
OX   NCBI_TaxID=9606;
RN   [1]
RP   NUCLEOTIDE SEQUENCE.
RC   TISSUE=Lung;
RX   MEDLINE=22388257; PubMed=12477932; DOI=10.1073/pnas.242603899;
RA   Strausberg R.L., Feingold E.A., Grouse L.H., Derge J.G.,
RA   Klausner R.D., Collins F.S., Wagner L., Shenmen C.M., Schuler G.D.,
RA   Altschul S.F., Zeeberg B., Buetow K.H., Schaefer C.F., Bhat N.K.,
RA   Hopkins R.F., Jordan H., Moore T., Max S.I., Wang J., Hsieh F.,
RA   Diatchenko L., Marusina K., Farmer A.A., Rubin G.M., Hong L.,
RA   Stapleton M., Soares M.B., Bonaldo M.F., Casavant T.L., Scheetz T.E.,
RA   Brownstein M.J., Usdin T.B., Toshiyuki S., Carninci P., Prange C.,
RA   Raha S.S., Loquellano N.A., Peters G.J., Abramson R.D., Mullahy S.J.,
RA   Bosak S.A., McEwan P.J., McKernan K.J., Malek J.A., Gunaratne P.H.,
RA   Richards S., Worley K.C., Hale S., Garcia A.M., Gay L.J., Hulyk S.W.,
RA   Villalon D.K., Muzny D.M., Sodergren E.J., Lu X., Gibbs R.A.,
RA   Fahey J., Helton E., Ketteman M., Madan A., Rodrigues S., Sanchez A.,
RA   Whiting M., Madan A., Young A.C., Shevchenko Y., Bouffard G.G.,
RA   Blakesley R.W., Touchman J.W., Green E.D., Dickson M.C.,
RA   Rodriguez A.C., Grimwood J., Schmutz J., Myers R.M.,
RA   Butterfield Y.S.N., Krzywinski M.I., Skalska U., Smailus D.E.,
RA   Schnerch A., Schein J.E., Jones S.J.M., Marra M.A.;
RT   "Generation and initial analysis of more than 15,000 full-length human
RT   and mouse cDNA sequences.";
RL   Proc. Natl. Acad. Sci. U.S.A. 99:16899-16903(2002).
RN   [2]
RP   NUCLEOTIDE SEQUENCE.
RC   TISSUE=Lung;
RA   Director MGC Project;
RL   Submitted (DEC-2002) to the EMBL/GenBank/DDBJ databases.
DR   EMBL; BC041709; AAH41709.1; -; mRNA.
DR   Ensembl; ENSG00000164463; Homo sapiens.
DR   InterPro; IPR004827; TF_bZIP.
DR   PROSITE; PS00036; BZIP_BASIC; UNKNOWN_1.
SQ   SEQUENCE   639 AA;  72118 MW;  ECFB92D9290DEDEB CRC64;

  Query Match             59.6%;  Score 381;  DB 2;  Length 639;
  Best Local Similarity   99.8%;  Pred. No. 0;
  Matches  481;  Conservative    0;  Mismatches    1;  Indels    0;  Gaps    0;

Qy          1 MPQPSVSGMDPPFGDAFRSHTFSEQTLMSTDLLANSSDPDFMYELDREMNYQQNPRDNFL 60
              ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db          1 MPQPSVSGMDPPFGDAFRSHTFSEQTLMSTDLLANSSDPDFMYELDREMNYQQNPRDNFL 60

Qy         61 SLEDCKDIENLESFTDVLDNEGALTSNWEQWDTYCEDLTKYTKLTSCDIWGTKEVDYLGL 120
              ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db         61 SLEDCKDIENLESFTDVLDNEGALTSNWEQWDTYCEDLTKYTKLTSCDIWGTKEVDYLGL 120

Qy        121 DDFSSPYQDEEVISKTPTLAQLNSEDSQSVSDSLYYPDSLFSVKQNPLPSSFPGKKITSR 180
              ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db        121 DDFSSPYQDEEVISKTPTLAQLNSEDSQSVSDSLYYPDSLFSVKQNPLPSSFPGKKITSR 180

Qy        181 AAAPVCSSKTLQAEVPLSDCVQKASKPPSSTQIMVKTNMYHNEKVNFHVECKDYVKKAKV 240
              |||||||||||||||||||||||||||||| ||||||||||||||||||||||||||||||
Db        181 AAAPVCSSKTLQAEVPLSDCVQKASKPTSSTQIMVKTNMYHNEKVNFHVECKDYVKKAKV 240

Qy        241 KINPVQQSRPLLSQIHTDAAKENTCYCGAVAKRQEKKGMEPLQGHATPALPFKETQELLL 300
              ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db        241 KINPVQQSRPLLSQIHTDAAKENTCYCGAVAKRQEKKGMEPLQGHATPALPFKETQELLL 300

Qy        301 SPLPQEGPGSLAAGESSSLSASTSVSDSSQKKEEHNYSLFVSDNLGEQPTKCSPEEDEED 360
              ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db        301 SPLPQEGPGSLAAGESSSLSASTSVSDSSQKKEEHNYSLFVSDNLGEQPTKCSPEEDEED 360

Qy        361 EEDVDDEDHDEGFGSEHELSENEEEEEEEEDYEDDKDDDISDTFSEPGYENDSVEDLKEV 420
              ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db        361 EEDVDDEDHDEGFGSEHELSENEEEEEEEEDYEDDKDDDISDTFSEPGYENDSVEDLKEV 420

Qy        421 TSISSRKRGKRRYFWEYSEQLTPSQQERMLRPSEWNRDTLPSNMYQKNGLHHGKYAVKKS 480
              ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db        421 TSISSRKRGKRRYFWEYSEQLTPSQQERMLRPSEWNRDTLPSNMYQKNGLHHGKYAVKKS 480

Qy        481 RR 482
              ||
Db        481 RR 482


RESULT 3
Q5HYG4_HUMAN
ID   Q5HYG4_HUMAN PRELIMINARY;      PRT;   408 AA.
AC   Q5HYG4;
DT   10-MAY-2005 (TrEMBLrel. 30, Created)
DT   10-MAY-2005 (TrEMBLrel. 30, Last sequence update)
DT   10-MAY-2005 (TrEMBLrel. 30, Last annotation update)
DE   Hypothetical protein DKFZp686G2059 (Fragment).
GN   Name=DKFZp686G2059;
OS   Homo sapiens (Human).
OC   Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;
OC   Mammalia; Eutheria; Euarchontoglires; Primates; Catarrhini; Hominidae;
OC   Homo.
OX   NCBI_TaxID=9606;
RN   [1]
RP   NUCLEOTIDE SEQUENCE.
RC   TISSUE=Endometrium;
RG   The German cDNA Consortium;
RA   Poustka A., Albert R., Moosmayer P., Schupp I., Wellenreuther R.,
RA   Mewes H.W., Weil B., Amid C., Osanger A., Fobo G., Han M., Wiemann S.;
RL   Submitted (JAN-2005) to the EMBL/GenBank/DDBJ databases.
DR   EMBL; BX647768; CAI46039.1; -; mRNA.
KW   Hypothetical protein.
FT   NON_TER     408    408
SQ   SEQUENCE   408 AA;  45625 MW;  E3E36BFEA8B4284B CRC64;

  Query Match             48.0%;  Score 307;  DB 2;  Length 408;
  Best Local Similarity   99.8%;  Pred. No. 1.1e-292;
  Matches  407;  Conservative    0;  Mismatches    1;  Indels    0;  Gaps    0;

Qy          1 MPQPSVSGMDPPFGDAFRSHTFSEQTLMSTDLLANSSDPDFMYELDREMNYQQNPRDNFL 60
              ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db          1 MPQPSVSGMDPPFGDAFRSHTFSEQTLMSTDLLANSSDPDFMYELDREMNYQQNPRDNFL 60

Qy         61 SLEDCKDIENLESFTDVLDNEGALTSNWEQWDTYCEDLTKYTKLTSCDIWGTKEVDYLGL 120
              ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db         61 SLEDCKDIENLESFTDVLDNEGALTSNWEQWDTYCEDLTKYTKLTSCDIWGTKEVDYLGL 120

Qy        121 DDFSSPYQDEEVISKTPTLAQLNSEDSQSVSDSLYYPDSLFSVKQNPLPSSFPGKKITSR 180
              ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db        121 DDFSSPYQDEEVISKTPTLAQLNSEDSQSVSDSLYYPDSLFSVKQNPLPSSFPGKKITSR 180

Qy        181 AAAPVCSSKTLQAEVPLSDCVQKASKPPSSTQIMVKTNMYHNEKVNFHVECKDYVKKAKV 240
              |||||||||||||||||||||||||||||| ||||||||||||||||||||||||||||||
Db        181 AAAPVCSSKTLQAEVPLSDCVQKASKPTSSTQIMVKTNMYHNEKVNFHVECKDYVKKAKV 240

Qy        241 KINPVQQSRPLLSQIHTDAAKENTCYCGAVAKRQEKKGMEPLQGHATPALPFKETQELLL 300
```

```
                ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db        241 KINPVQQSRPLLSQIHTDAAKENTCYCGAVAKRQEKKGMEPLQGHATPALPFKETQELLL 300

Qy        301 SPLPQEGPGSLAAGESSSLSASTSVSDSSQKKEEHNYSLFVSDNLGEQPTKCSPEEDEED 360
                ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db        301 SPLPQEGPGSLAAGESSSLSASTSVSDSSQKKEEHNYSLFVSDNLGEQPTKCSPEEDEED 360

Qy        361 EEDVDDEDHDEGFGSEHELSENEEEEEEEEDYEDDKDDDISDTFSEPG 408
                |||||||||||||||||||||||||||||||||||||||||||||||||
Db        361 EEDVDDEDHDEGFGSEHELSENEEEEEEEEDYEDDKDDDISDTFSEPG 408


RESULT 4
Q86YR3_HUMAN
ID   Q86YR3_HUMAN PRELIMINARY;      PRT;   417 AA.
AC   Q86YR3;
DT   01-JUN-2003 (TrEMBLrel. 24, Created)
DT   01-JUN-2003 (TrEMBLrel. 24, Last sequence update)
DT   01-JUN-2003 (TrEMBLrel. 24, Last annotation update)
DE   Hypothetical protein.
OS   Homo sapiens (Human).
OC   Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;
OC   Mammalia; Eutheria; Euarchontoglires; Primates; Catarrhini; Hominidae;
OC   Homo.
OX   NCBI_TaxID=9606;
RN   [1]
RP   NUCLEOTIDE SEQUENCE.
RA   Li F., Yao K.T.;
RL   Submitted (NOV-2002) to the EMBL/GenBank/DDBJ databases.
DR   EMBL; AY174896; AAO18732.1; -; mRNA.
KW   Hypothetical protein.
SQ   SEQUENCE   417 AA;  46579 MW;  9DBD37B07C14556B CRC64;

  Query Match             48.0%;  Score 307;  DB 2;  Length 417;
  Best Local Similarity   99.8%;  Pred. No. 1.1e-292;
  Matches  407;  Conservative    0;  Mismatches    1;  Indels    0;  Gaps    0;

Qy          1 MPQPSVSGMDPPFGDAFRSHTFSEQTLMSTDLLANSSDPDFMYELDREMNYQQNPRDNFL 60
              ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db          1 MPQPSVSGMDPPFGDAFRSHTFSEQTLMSTDLLANSSDPDFMYELDREMNYQQNPRDNFL 60

Qy         61 SLEDCKDIENLESFTDVLDNEGALTSNWEQWDTYCEDLTKYTKLTSCDIWGTKEVDYLGL 120
              ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db         61 SLEDCKDIENLESFTDVLDNEGALTSNWEQWDTYCEDLTKYTKLTSCDIWGTKEVDYLGL 120

Qy        121 DDFSSPYQDEEVISKTPTLAQLNSEDSQSVSDSLYYPDSLFSVKQNPLPSSFPGKKITSR 180
              ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db        121 DDFSSPYQDEEVISKTPTLAQLNSEDSQSVSDSLYYPDSLFSVKQNPLPSSFPGKKITSR 180

Qy        181 AAAPVCSSKTLQAEVPLSDCVQKASKPPSSTQIMVKTNMYHNEKVNFHVECKDYVKKAKV 240
              |||||||||||||||||||||| ||||||||||||||||||||||||||||||||||||||||
Db        181 AAAPVCSSKTLQAEVPLSDCVQKASKPTSSTQIMVKTNMYHNEKVNFHVECKDYVKKAKV 240

Qy        241 KINPVQQSRPLLSQIHTDAAKENTCYCGAVAKRQEKKGMEPLQGHATPALPFKETQELLL 300
              ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db        241 KINPVQQSRPLLSQIHTDAAKENTCYCGAVAKRQEKKGMEPLQGHATPALPFKETQELLL 300

Qy        301 SPLPQEGPGSLAAGESSSLSASTSVSDSSQKKEEHNYSLFVSDNLGEQPTKCSPEEDEED 360
              ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db        301 SPLPQEGPGSLAAGESSSLSASTSVSDSSQKKEEHNYSLFVSDNLGEQPTKCSPEEDEED 360

Qy        361 EEDVDDEDHDEGFGSEHELSENEEEEEEEEDYEDDKDDDISDTFSEPG 408
              |||||||||||||||||||||||||||||||||||||||||||||||||
Db        361 EEDVDDEDHDEGFGSEHELSENEEEEEEEEDYEDDKDDDISDTFSEPG 408


RESULT 5
Q5HYK0_HUMAN
ID   Q5HYK0_HUMAN PRELIMINARY;      PRT;   604 AA.
AC   Q5HYK0;
DT   10-MAY-2005 (TrEMBLrel. 30, Created)
DT   10-MAY-2005 (TrEMBLrel. 30, Last sequence update)
DT   10-MAY-2005 (TrEMBLrel. 30, Last annotation update)
DE   Hypothetical protein DKFZp313F2319 (Fragment).
GN   Name=DKFZp313F2319;
OS   Homo sapiens (Human).
OC   Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;
OC   Mammalia; Eutheria; Euarchontoglires; Primates; Catarrhini; Hominidae;
OC   Homo.
OX   NCBI_TaxID=9606;
RN   [1]
RP   NUCLEOTIDE SEQUENCE.
RC   TISSUE=Adipose;
RG   The German cDNA Consortium;
RA   Koehrer K., Beyer A., Mewes H.W., Weil B., Amid C., Osanger A.,
RA   Fobo G., Han M., Wiemann S.;
RL   Submitted (JAN-2005) to the EMBL/GenBank/DDBJ databases.
DR   EMBL; BX647573; CAI46104.1; -; mRNA.
DR   InterPro; IPR004827; TF_bZIP.
DR   PROSITE; PS00036; BZIP_BASIC; UNKNOWN_1.
KW   Hypothetical protein.
FT   NON_TER     604     604
SQ   SEQUENCE   604 AA;  68572 MW;  977C229B63E2E4C2 CRC64;

  Query Match             42.6%;  Score 272;  DB 2;  Length 604;
  Best Local Similarity   99.6%;  Pred. No. 4.2e-258;
  Matches  472;  Conservative    0;  Mismatches    2;  Indels    0;  Gaps    0;

Qy          1 MPQPSVSGMDPPFGDAFRSHTFSEQTLMSTDLLANSSDPDFMYELDREMNYQQNPRDNFL 60
              ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db          1 MPQPSVSGMDPPFGDAFRSHTFSEQTLMSTDLLANSSDPDFMYELDREMNYQQNPRDNFL 60

Qy         61 SLEDCKDIENLESFTDVLDNEGALTSNWEQWDTYCEDLTKYTKLTSCDIWGTKEVDYLGL 120
              ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db         61 SLEDCKDIENLESFTDVLDNEGALTSNWEQWDTYCEDLTKYTKLTSCDIWGTKEVDYLGL 120

Qy        121 DDFSSPYQDEEVISKTPTLAQLNSEDSQSVSDSLYYPDSLFSVKQNPLPSSFPGKKITSR 180
              ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db        121 DDFSSPYQDEEVISKTPTLAQLNSEDSQSVSDSLYYPDSLFSVKQNPLPSSFPGKKITSR 180

Qy        181 AAAPVCSSKTLQAEVPLSDCVQKASKPPSSTQIMVKTNMYHNEKVNFHVECKDYVKKAKV 240
              |||||||||||||||||||||| ||||||||||||||||||||||||||||||||||||||||
Db        181 AAAPVCSSKTLQAEVPLSDCVQKASKPTSSTQIMVKTNMYHNEKVNFHVECRDYVKKAKV 240

Qy        241 KINPVQQSRPLLSQIHTDAAKENTCYCGAVAKRQEKKGMEPLQGHATPALPFKETQELLL 300
              ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db        241 KINPVQQSRPLLSQIHTDAAKENTCYCGAVAKRQEKKGMEPLQGHATPALPFKETQELLL 300

Qy        301 SPLPQEGPGSLAAGESSSLSASTSVSDSSQKKEEHNYSLFVSDNLGEQPTKCSPEEDEED 360
              ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db        301 SPLPQEGPGSLAAGESSSLSASTSVSDSSQKKEEHNYSLFVSDNLGEQPTKCSPEEDEED 360

Qy        361 EEDVDDEDHDEGFGSEHELSENEEEEEEEEDYEDDKDDDISDTFSEPGYENDSVEDLKEV 420
              ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db        361 EEDVDDEDHDEGFGSEHELSENEEEEEEEEDYEDDKDDDISDTFSEPGYENDSVEDLKEV 420

Qy        421 TSISSRKRGKRRYFWEYSEQLTPSQQERMLRPSEWNRDTLPSNMYQKNGLHHGK 474
              ||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db        421 TSISSRKRGKRRYFWEYSEQLTPSQQERMLRPSEWNRDTLPSNMYQKNGLHHGK 474


RESULT 6
Q9CTQ7_MOUSE
ID   Q9CTQ7_MOUSE PRELIMINARY;      PRT;   330 AA.
AC   Q9CTQ7;
DT   01-JUN-2001 (TrEMBLrel. 17, Created)
DT   01-JUN-2001 (TrEMBLrel. 17, Last sequence update)
DT   01-JUN-2001 (TrEMBLrel. 17, Last annotation update)
DE   Mus musculus adult retina cDNA, RIKEN full-length enriched library,
DE   clone:A930001N09 product:hypothetical protein, full insert sequence.
DE   (Fragment).
GN   Name=A930001N09Rik;
OS   Mus musculus (Mouse).
OC   Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;
OC   Mammalia; Eutheria; Euarchontoglires; Glires; Rodentia; Sciurognathi;
OC   Muroidea; Muridae; Murinae; Mus.
```

```
OX   NCBI_TaxID=10090;
RN   [1]
RP   NUCLEOTIDE SEQUENCE.
RC   STRAIN=C57BL/6J; TISSUE=Retina;
RX   MEDLINE=99279253; PubMed=10349636; DOI=10.1016/S0076-6879(99)03004-9;
RA   Carninci P., Hayashizaki Y.;
RT   "High-efficiency full-length cDNA cloning.";
RL   Meth. Enzymol. 303:19-44(1999).
RN   [2]
RP   NUCLEOTIDE SEQUENCE.
RC   STRAIN=C57BL/6J; TISSUE=Retina;
RX   MEDLINE=21085660; PubMed=11217851; DOI=10.1038/35055500;
RA   Kawai J., Shinagawa A., Shibata K., Yoshino M., Itoh M., Ishii Y.,
RA   Arakawa T., Hara A., Fukunishi Y., Konno H., Adachi J., Fukuda S.,
RA   Aizawa K., Izawa M., Nishi K., Kiyosawa H., Kondo S., Yamanaka I.,
RA   Saito T., Okazaki Y., Gojobori T., Bono H., Kasukawa T., Saito R.,
RA   Kadota K., Matsuda H.A., Ashburner M., Batalov S., Casavant T.,
RA   Fleischmann W., Gaasterland T., Gissi C., King B., Kochiwa H.,
RA   Kuehl P., Lewis S., Matsuo Y., Nikaido I., Pesole G., Quackenbush J.,
RA   Schriml L.M., Staubli F., Suzuki R., Tomita M., Wagner L., Washio T.,
RA   Sakai K., Okido T., Furuno M., Aono H., Baldarelli R., Barsh G.,
RA   Blake J., Boffelli D., Bojunga N., Carninci P., de Bonaldo M.F.,
RA   Brownstein M.J., Bult C., Fletcher C., Fujita M., Gariboldi M.,
RA   Gustincich S., Hill D., Hofmann M., Hume D.A., Kamiya M., Lee N.H.,
RA   Lyons P., Marchionni L., Mashima J., Mazzarelli J., Mombaerts P.,
RA   Nordone P., Ring B., Ringwald M., Rodriguez I., Sakamoto N.,
RA   Sasaki H., Sato K., Schoenbach C., Seya T., Shibata Y., Storch K.-F.,
RA   Suzuki H., Toyo-oka K., Wang K.H., Weitz C., Whittaker C., Wilming L.,
RA   Wynshaw-Boris A., Yoshida K., Hasegawa Y., Kawaji H., Kohtsuki S.,
RA   Hayashizaki Y.;
RT   "Functional annotation of a full-length mouse cDNA collection.";
RL   Nature 409:685-690(2001).
RN   [3]
RP   NUCLEOTIDE SEQUENCE.
RC   STRAIN=C57BL/6J; TISSUE=Retina;
RA   The FANTOM Consortium,
RA   the RIKEN Genome Exploration Research Group Phase I & II Team;
RT   "Analysis of the mouse transcriptome based on functional annotation of
RT   60,770 full-length cDNAs.";
RL   Nature 420:563-573(2002).
RN   [4]
RP   NUCLEOTIDE SEQUENCE.
RC   STRAIN=C57BL/6J; TISSUE=Retina;
RX   MEDLINE=20499374; PubMed=11042159; DOI=10.1101/gr.145100;
RA   Carninci P., Shibata Y., Hayatsu N., Sugahara Y., Shibata K., Itoh M.,
RA   Konno H., Okazaki Y., Muramatsu M., Hayashizaki Y.;
RT   "Normalization and subtraction of cap-trapper-selected cDNAs to
RT   prepare full-length cDNA libraries for rapid discovery of new genes.";
RL   Genome Res. 10:1617-1630(2000).
RN   [5]
RP   NUCLEOTIDE SEQUENCE.
RC   STRAIN=C57BL/6J; TISSUE=Retina;
RX   MEDLINE=20530913; PubMed=11076861; DOI=10.1101/gr.152600;
RA   Shibata K., Itoh M., Aizawa K., Nagaoka S., Sasaki N., Carninci P.,
RA   Konno H., Akiyama J., Nishi K., Kitsunai T., Tashiro H., Itoh M.,
RA   Sumi N., Ishii Y., Nakamura S., Hazama M., Nishine T., Harada A.,
RA   Yamamoto R., Matsumoto H., Sakaguchi S., Ikegami T., Kashiwagi K.,
RA   Fujiwake S., Inoue K., Togawa Y., Izawa M., Ohara E., Watahiki M.,
RA   Yoneda Y., Ishikawa T., Ozawa K., Tanaka T., Matsuura S., Kawai J.,
RA   Okazaki Y., Muramatsu M., Inoue Y., Kira A., Hayashizaki Y.;
RT   "RIKEN integrated sequence analysis (RISA) system-384-format
RT   sequencing pipeline with 384 multicapillary sequencer.";
RL   Genome Res. 10:1757-1771(2000).
RN   [6]
RP   NUCLEOTIDE SEQUENCE.
RC   STRAIN=C57BL/6J; TISSUE=Retina;
RA   Adachi J., Aizawa K., Akahira S., Akimura T., Arai A., Aono H.,
RA   Arakawa T., Bono H., Carninci P., Fukuda S., Fukunishi Y., Furuno M.,
RA   Hanagaki T., Hara A., Hayatsu N., Hiramoto K., Hiraoka T., Hori F.,
RA   Imotani K., Ishii Y., Itoh M., Izawa M., Kasukawa T., Kato H.,
RA   Kawai J., Kojima Y., Konno H., Kouda M., Koya S., Kurihara C.,
RA   Matsuyama T., Miyazaki A., Nishi K., Nomura K., Numazaki R., Ohno M.,
RA   Okazaki Y., Okido T., Owa C., Saito H., Saito R., Sakai C., Sakai K.,
RA   Sano H., Sasaki D., Shibata K., Shibata Y., Shinagawa A., Shiraki T.,
RA   Sogabe Y., Suzuki H., Tagami M., Tagawa A., Takahashi F., Tanaka T.,
RA   Tejima Y., Toya T., Yamamura T., Yasunishi A., Yoshida K., Yoshino M.,
RA   Muramatsu M., Hayashizaki Y.;
RL   Submitted (AUG-2000) to the EMBL/GenBank/DDBJ databases.
DR   EMBL; AK020796; BAB32214.1; -; mRNA.
DR   Ensembl; ENSMUSG00000048249; Mus musculus.
DR   MGI; MGI:1924378; A930001N09Rik.
KW   Hypothetical protein.
FT   NON_TER     330    330
SQ   SEQUENCE   330 AA;  36692 MW;  1397555C4934A64B CRC64;

  Query Match             13.6%;  Score 87;  DB 2;  Length 330;
  Best Local Similarity   100.0%;  Pred. No. 2.8e-76;
  Matches   87;  Conservative    0;  Mismatches    0;  Indels    0;  Gaps    0;

Qy         83 ALTSNWEQWDTYCEDLTKYTKLTSCDIWGTKEVDYLGLDDFSSPYQDEEVISKTPTLAQL 142
              ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db         83 ALTSNWEQWDTYCEDLTKYTKLTSCDIWGTKEVDYLGLDDFSSPYQDEEVISKTPTLAQL 142

Qy        143 NSEDSQSVSDSLYYPDSLFSVKQNPLP 169
              ||||||||||||||||||||||||||||
Db        143 NSEDSQSVSDSLYYPDSLFSVKQNPLP 169


RESULT 7
Q8C8N7_MOUSE
ID   Q8C8N7_MOUSE PRELIMINARY;      PRT;   351 AA.
AC   Q8C8N7;
DT   01-MAR-2003 (TrEMBLrel. 23, Created)
DT   01-MAR-2003 (TrEMBLrel. 23, Last sequence update)
DT   01-JUN-2003 (TrEMBLrel. 24, Last annotation update)
DE   Mus musculus adult retina cDNA, RIKEN full-length enriched library,
DE   clone:A930040G19 product:hypothetical protein, full insert sequence.
DE   (Fragment).
GN   Name=A930001N09Rik;
OS   Mus musculus (Mouse).
OC   Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;
OC   Mammalia; Eutheria; Euarchontoglires; Glires; Rodentia; Sciurognathi;
OC   Muroidea; Muridae; Murinae; Mus.
OX   NCBI_TaxID=10090;
RN   [1]
RP   NUCLEOTIDE SEQUENCE.
RC   STRAIN=C57BL/6J; TISSUE=Retina;
RX   MEDLINE=99279253; PubMed=10349636; DOI=10.1016/S0076-6879(99)03004-9;
RA   Carninci P., Hayashizaki Y.;
RT   "High-efficiency full-length cDNA cloning.";
RL   Meth. Enzymol. 303:19-44(1999).
RN   [2]
RP   NUCLEOTIDE SEQUENCE.
RC   STRAIN=C57BL/6J; TISSUE=Retina;
RX   MEDLINE=21085660; PubMed=11217851; DOI=10.1038/35055500;
RA   Kawai J., Shinagawa A., Shibata K., Yoshino M., Itoh M., Ishii Y.,
RA   Arakawa T., Hara A., Fukunishi Y., Konno H., Adachi J., Fukuda S.,
RA   Aizawa K., Izawa M., Nishi K., Kiyosawa H., Kondo S., Yamanaka I.,
RA   Saito T., Okazaki Y., Gojobori T., Bono H., Kasukawa T., Saito R.,
RA   Kadota K., Matsuda H.A., Ashburner M., Batalov S., Casavant T.,
RA   Fleischmann W., Gaasterland T., Gissi C., King B., Kochiwa H.,
RA   Kuehl P., Lewis S., Matsuo Y., Nikaido I., Pesole G., Quackenbush J.,
RA   Schriml L.M., Staubli F., Suzuki R., Tomita M., Wagner L., Washio T.,
RA   Sakai K., Okido T., Furuno M., Aono H., Baldarelli R., Barsh G.,
RA   Blake J., Boffelli D., Bojunga N., Carninci P., de Bonaldo M.F.,
RA   Brownstein M.J., Bult C., Fletcher C., Fujita M., Gariboldi M.,
RA   Gustincich S., Hill D., Hofmann M., Hume D.A., Kamiya M., Lee N.H.,
RA   Lyons P., Marchionni L., Mashima J., Mazzarelli J., Mombaerts P.,
RA   Nordone P., Ring B., Ringwald M., Rodriguez I., Sakamoto N.,
RA   Sasaki H., Sato K., Schoenbach C., Seya T., Shibata Y., Storch K.-F.,
RA   Suzuki H., Toyo-oka K., Wang K.H., Weitz C., Whittaker C., Wilming L.,
RA   Wynshaw-Boris A., Yoshida K., Hasegawa Y., Kawaji H., Kohtsuki S.,
RA   Hayashizaki Y.;
RT   "Functional annotation of a full-length mouse cDNA collection.";
RL   Nature 409:685-690(2001).
RN   [3]
```

```
RP   NUCLEOTIDE SEQUENCE.
RC   STRAIN=C57BL/6J; TISSUE=Retina;
RA   The FANTOM Consortium,
RA   the RIKEN Genome Exploration Research Group Phase I & II Team;
RT   "Analysis of the mouse transcriptome based on functional annotation of
RT   60,770 full-length cDNAs.";
RL   Nature 420:563-573(2002).
RN   [4]
RP   NUCLEOTIDE SEQUENCE.
RC   STRAIN=C57BL/6J; TISSUE=Retina;
RX   MEDLINE=20499374; PubMed=11042159; DOI=10.1101/gr.145100;
RA   Carninci P., Shibata Y., Hayatsu N., Sugahara Y., Shibata K., Itoh M.,
RA   Konno H., Okazaki Y., Muramatsu M., Hayashizaki Y.;
RT   "Normalization and subtraction of cap-trapper-selected cDNAs to
RT   prepare full-length cDNA libraries for rapid discovery of new genes.";
RL   Genome Res. 10:1617-1630(2000).
RN   [5]
RP   NUCLEOTIDE SEQUENCE.
RC   STRAIN=C57BL/6J; TISSUE=Retina;
RX   MEDLINE=20530913; PubMed=11076861; DOI=10.1101/gr.152600;
RA   Shibata K., Itoh M., Aizawa K., Nagaoka S., Sasaki N., Carninci P.,
RA   Konno H., Akiyama J., Nishi K., Kitsunai T., Tashiro H., Itoh M.,
RA   Sumi N., Ishii Y., Nakamura S., Hazama M., Nishine T., Harada A.,
RA   Yamamoto R., Matsumoto H., Sakaguchi S., Ikegami T., Kashiwagi K.,
RA   Fujiwake S., Inoue K., Togawa Y., Izawa M., Ohara E., Watahiki M.,
RA   Yoneda Y., Ishikawa T., Ozawa K., Tanaka T., Matsuura S., Kawai J.,
RA   Okazaki Y., Muramatsu M., Inoue Y., Kira A., Hayashizaki Y.;
RT   "RIKEN integrated sequence analysis (RISA) system-384-format
RT   sequencing pipeline with 384 multicapillary sequencer.";
RL   Genome Res. 10:1757-1771(2000).
RN   [6]
RP   NUCLEOTIDE SEQUENCE.
RC   STRAIN=C57BL/6J; TISSUE=Retina;
RA   Adachi J., Aizawa K., Akimura T., Arakawa T., Bono H., Carninci P.,
RA   Fukuda S., Furuno M., Hanagaki T., Hara A., Hashizume W.,
RA   Hayashida K., Hayatsu N., Hiramoto K., Hiraoka T., Hirozane T.,
RA   Hori F., Imotani K., Ishii Y., Itoh M., Kagawa I., Kasukawa T.,
RA   Katoh H., Kawai J., Kojima Y., Kondo S., Konno H., Kouda M., Koya S.,
RA   Kurihara C., Matsuyama T., Miyazaki A., Murata M., Nakamura M.,
RA   Nishi K., Nomura K., Numazaki R., Ohno M., Ohsato N., Okazaki Y.,
RA   Saito R., Saitoh H., Sakai C., Sakai K., Sakazume N., Sano H.,
RA   Sasaki D., Shibata K., Shinagawa A., Shiraki T., Sogabe Y., Tagami M.,
RA   Tagawa A., Takahashi F., Takaku-Akahira S., Takeda Y., Tanaka T.,
RA   Tomaru A., Toya T., Yasunishi A., Muramatsu M., Hayashizaki Y.;
RL   Submitted (JUL-2001) to the EMBL/GenBank/DDBJ databases.
DR   EMBL; AK044760; BAC32070.1; -; mRNA.
DR   Ensembl; ENSMUSG00000048249; Mus musculus.
DR   MGI; MGI:1924378; A930001N09Rik.
KW   Hypothetical protein.
FT   NON_TER     351    351
SQ   SEQUENCE   351 AA;  39254 MW;  ABA4B36D996AFAAD CRC64;

  Query Match             13.6%;  Score 87;  DB 2;  Length 351;
  Best Local Similarity   100.0%;  Pred. No. 2.9e-76;
  Matches   87;  Conservative    0;  Mismatches    0;  Indels    0;  Gaps    0;

Qy         83 ALTSNWEQWDTYCEDLTKYTKLTSCDIWGTKEVDYLGLDDFSSPYQDEEVISKTPTLAQL 142
              |||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db         83 ALTSNWEQWDTYCEDLTKYTKLTSCDIWGTKEVDYLGLDDFSSPYQDEEVISKTPTLAQL 142

Qy        143 NSEDSQSVSDSLYYPDSLFSVKQNPLP 169
              ||||||||||||||||||||||||||
Db        143 NSEDSQSVSDSLYYPDSLFSVKQNPLP 169


RESULT 8
Q8CDG5_MOUSE
ID   Q8CDG5_MOUSE PRELIMINARY;      PRT;   640 AA.
AC   Q8CDG5;
DT   01-MAR-2003 (TrEMBLrel. 23, Created)
DT   01-MAR-2003 (TrEMBLrel. 23, Last sequence update)
DT   01-MAR-2004 (TrEMBLrel. 26, Last annotation update)
DE   Mus musculus adult male testis cDNA, RIKEN full-length enriched
DE   library, clone:4932441F15 product:hypothetical protein, full insert
DE   sequence.
GN   Name=A930001N09Rik;
OS   Mus musculus (Mouse).
OC   Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;
OC   Mammalia; Eutheria; Euarchontoglires; Glires; Rodentia; Sciurognathi;
OC   Muroidea; Muridae; Murinae; Mus.
OX   NCBI_TaxID=10090;
RN   [1]
RP   NUCLEOTIDE SEQUENCE.
RC   STRAIN=C57BL/6J; TISSUE=Testis;
RX   MEDLINE=99279253; PubMed=10349636; DOI=10.1016/S0076-6879(99)03004-9;
RA   Carninci P., Hayashizaki Y.;
RT   "High-efficiency full-length cDNA cloning.";
RL   Meth. Enzymol. 303:19-44(1999).
RN   [2]
RP   NUCLEOTIDE SEQUENCE.
RC   STRAIN=C57BL/6J; TISSUE=Testis;
RX   MEDLINE=21085660; PubMed=11217851; DOI=10.1038/35055500;
RA   Kawai J., Shinagawa A., Shibata K., Yoshino M., Itoh M., Ishii Y.,
RA   Arakawa T., Hara A., Fukunishi Y., Konno H., Adachi J., Fukuda S.,
RA   Aizawa K., Izawa M., Nishi K., Kiyosawa H., Kondo S., Yamanaka I.,
RA   Saito T., Okazaki Y., Gojobori T., Bono H., Kasukawa T., Saito R.,
RA   Kadota K., Matsuda H.A., Ashburner M., Batalov S., Casavant T.,
RA   Fleischmann W., Gaasterland T., Gissi C., King B., Kochiwa H.,
RA   Kuehl P., Lewis S., Matsuo Y., Nikaido I., Pesole G., Quackenbush J.,
RA   Schriml L.M., Staubli F., Suzuki R., Tomita M., Wagner L., Washio T.,
RA   Sakai K., Okido T., Furuno M., Aono H., Baldarelli R., Barsh G.,
RA   Blake J., Boffelli D., Bojunga N., Carninci P., de Bonaldo M.F.,
RA   Brownstein M.J., Bult C., Fletcher C., Fujita M., Gariboldi M.,
RA   Gustincich S., Hill D., Hofmann M., Hume D.A., Kamiya M., Lee N.H.,
RA   Lyons P., Marchionni L., Mashima J., Mazzarelli J., Mombaerts P.,
RA   Nordone P., Ring B., Ringwald M., Rodriguez I., Sakamoto N.,
RA   Sasaki H., Sato K., Schoenbach C., Seya T., Shibata Y., Storch K.-F.,
RA   Suzuki H., Toyo-oka K., Wang K.H., Weitz C., Whittaker C., Wilming L.,
RA   Wynshaw-Boris A., Yoshida K., Hasegawa Y., Kawaji H., Kohtsuki S.,
RA   Hayashizaki Y.;
RT   "Functional annotation of a full-length mouse cDNA collection.";
RL   Nature 409:685-690(2001).
RN   [3]
RP   NUCLEOTIDE SEQUENCE.
RC   STRAIN=C57BL/6J; TISSUE=Testis;
RA   The FANTOM Consortium,
RA   the RIKEN Genome Exploration Research Group Phase I & II Team;
RT   "Analysis of the mouse transcriptome based on functional annotation of
RT   60,770 full-length cDNAs.";
RL   Nature 420:563-573(2002).
RN   [4]
RP   NUCLEOTIDE SEQUENCE.
RC   STRAIN=C57BL/6J; TISSUE=Testis;
RX   MEDLINE=20499374; PubMed=11042159; DOI=10.1101/gr.145100;
RA   Carninci P., Shibata Y., Hayatsu N., Sugahara Y., Shibata K., Itoh M.,
RA   Konno H., Okazaki Y., Muramatsu M., Hayashizaki Y.;
RT   "Normalization and subtraction of cap-trapper-selected cDNAs to
RT   prepare full-length cDNA libraries for rapid discovery of new genes.";
RL   Genome Res. 10:1617-1630(2000).
RN   [5]
RP   NUCLEOTIDE SEQUENCE.
RC   STRAIN=C57BL/6J; TISSUE=Testis;
RX   MEDLINE=20530913; PubMed=11076861; DOI=10.1101/gr.152600;
RA   Shibata K., Itoh M., Aizawa K., Nagaoka S., Sasaki N., Carninci P.,
RA   Konno H., Akiyama J., Nishi K., Kitsunai T., Tashiro H., Itoh M.,
RA   Sumi N., Ishii Y., Nakamura S., Hazama M., Nishine T., Harada A.,
RA   Yamamoto R., Matsumoto H., Sakaguchi S., Ikegami T., Kashiwagi K.,
RA   Fujiwake S., Inoue K., Togawa Y., Izawa M., Ohara E., Watahiki M.,
RA   Yoneda Y., Ishikawa T., Ozawa K., Tanaka T., Matsuura S., Kawai J.,
RA   Okazaki Y., Muramatsu M., Inoue Y.; Kira A., Hayashizaki Y.;
RT   "RIKEN integrated sequence analysis (RISA) system-384-format
RT   sequencing pipeline with 384 multicapillary sequencer.";
RL   Genome Res. 10:1757-1771(2000).
RN   [6]
RP   NUCLEOTIDE SEQUENCE.
RC   STRAIN=C57BL/6J; TISSUE=Testis;
```

```
RA   Adachi J., Aizawa K., Akimura T., Arakawa T., Bono H., Carninci P.,
RA   Fukuda S., Furuno M., Hanagaki T., Hara A., Hashizume W.,
RA   Hayashida K., Hayatsu N., Hiramoto K., Hiraoka T., Hirozane T.,
RA   Hori F., Imotani K., Ishii Y., Itoh M., Kagawa I., Kasukawa T.,
RA   Katoh H., Kawai J., Kojima Y., Kondo S., Konno H., Kouda M., Koya S.,
RA   Kurihara C., Matsuyama T., Miyazaki A., Murata M., Nakamura M.,
RA   Nishi K., Nomura K., Numazaki R., Ohno M., Ohsato N., Okazaki Y.,
RA   Saito R., Saitoh H., Sakai C., Sakai K., Sakazume N., Sano H.,
RA   Sasaki D., Shibata K., Shinagawa A., Shiraki T., Sogabe Y., Tagami M.,
RA   Tagawa A., Takahashi F., Takaku-Akahira S., Takeda Y., Tanaka T.,
RA   Tomaru A., Toya T., Yasunishi A., Muramatsu M., Hayashizaki Y.;
RL   Submitted (JUL-2001) to the EMBL/GenBank/DDBJ databases.
DR   EMBL; AK030092; BAC26779.1; -; mRNA.
DR   Ensembl; ENSMUSG00000048249; Mus musculus.
DR   MGI; MGI:1924378; A930001N09Rik.
DR   InterPro; IPR004827; TF_bZIP.
DR   PROSITE; PS00036; BZIP_BASIC; UNKNOWN_1.
KW   Hypothetical protein.
SQ   SEQUENCE   640 AA;  72598 MW;  FE02C532FA34E1DE CRC64;

  Query Match             13.6%;  Score 87;  DB 2;  Length 640;
  Best Local Similarity   100.0%;  Pred. No. 4.9e-76;
  Matches   87;  Conservative    0;  Mismatches    0;  Indels    0;  Gaps    0;

Qy         83 ALTSNWEQWDTYCEDLTKYTKLTSCDIWGTKEVDYLGLDDFSSPYQDEEVISKTPTLAQL 142
              ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db         83 ALTSNWEQWDTYCEDLTKYTKLTSCDIWGTKEVDYLGLDDFSSPYQDEEVISKTPTLAQL 142

Qy        143 NSEDSQSVSDSLYYPDSLFSVKQNPLP 169
              ||||||||||||||||||||||||||||
Db        143 NSEDSQSVSDSLYYPDSLFSVKQNPLP 169


RESULT 9
Q5ZMD5_CHICK
ID   Q5ZMD5_CHICK PRELIMINARY;      PRT;   131 AA.
AC   Q5ZMD5;
DT   25-OCT-2004 (TrEMBLrel. 28, Created)
DT   25-OCT-2004 (TrEMBLrel. 28, Last sequence update)
DT   25-OCT-2004 (TrEMBLrel. 28, Last annotation update)
DE   Hypothetical protein.
GN   ORFNames=RCJMB04_2h14;
OS   Gallus gallus (Chicken).
OC   Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;
OC   Archosauria; Aves; Neognathae; Galliformes; Phasianidae; Phasianinae;
OC   Gallus.
OX   NCBI_TaxID=9031;
RN   [1]
RP   NUCLEOTIDE SEQUENCE.
RC   STRAIN=CB; TISSUE=Bursa;
RA   Caldwell R.B., Kierzek A.M., Arakawa H., Bezzubov Y., Zaim J.,
RA   Fiedler P., Kutter S., Blagodatski A., Kostovska D., Koter M.,
RA   Plachy J., Carninci P., Hayashizaki Y., Buerstedde J.M.;
RT   "Full-length cDNAs from chicken bursal lymphocytes to facilitate
RT   genefunction analysis.";
RL   Genome Biol. 6:R6-R6(2005).
DR   EMBL; AJ719449; CAG31108.1; -; mRNA.
KW   Hypothetical protein.
SQ   SEQUENCE   131 AA;  15183 MW;  4D73D84C28DE3670 CRC64;

  Query Match              6.9%;  Score 44;  DB 2;  Length 131;
  Best Local Similarity   100.0%;  Pred. No. 2.5e-34;
  Matches   44;  Conservative    0;  Mismatches    0;  Indels    0;  Gaps    0;

Qy         88 WEQWDTYCEDLTKYTKLTSCDIWGTKEVDYLGLDDFSSPYQDEE 131
              |||||||||||||||||||||||||||||||||||||||||||||
Db         88 WEQWDTYCEDLTKYTKLTSCDIWGTKEVDYLGLDDFSSPYQDEE 131


RESULT 10
Q4RRX3_TETNG
ID   Q4RRX3_TETNG PRELIMINARY;      PRT;   600 AA.
AC   Q4RRX3;
DT   13-SEP-2005 (TrEMBLrel. 31, Created)
DT   13-SEP-2005 (TrEMBLrel. 31, Last sequence update)
DT   13-SEP-2005 (TrEMBLrel. 31, Last annotation update)
DE   Chromosome 7 SCAF15001, whole genome shotgun sequence.
DE   (Fragment).
GN   ORFNames=GSTENG00029962001;
OS   Tetraodon nigroviridis (Green puffer).
OC   Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;
OC   Actinopterygii; Neopterygii; Teleostei; Euteleostei; Neoteleostei;
OC   Acanthomorpha; Acanthopterygii; Percomorpha; Tetraodontiformes;
OC   Tetradontoidea; Tetraodontidae; Tetraodon.
OX   NCBI_TaxID=99883;
RN   [1]
RP   NUCLEOTIDE SEQUENCE.
RA   Jaillon O., Aury J.M., Brunet F., Petit J.L., Stange-Thomann N.,
RA   Mauceli E., Bouneau L., Fischer C., Ozouf-Costaz C., Bernot A.,
RA   Nicaud S., Jaffe D., Fisher S., Lutfalla G., Dossat C., Segurens B.,
RA   Dasilva C., Salanoubat M., Levy M., Boudet N., Castellano S.,
RA   Anthouard V., Jubin C., Castelli V., Katinka M., Vacherie B.,
RA   Biemont C., Skalli Z., Cattolico L., Poulain J., De Berardinis V.,
RA   Cruaud C., Duprat S., Brottier P., Coutanceau J.P., Gouzy J.,
RA   Parra G., Lardier G., Chapple C., McKernan K.J., McEwan P., Bosak S.,
RA   Kellis M., Volff JN., Guigo R., Zody M.C., Mesirov J.,
RA   Lindblad-Toh K., Birren B., Nusbaum C., Kahn D., Robinson-Rechavi M.,
RA   Laudet V., Schachter V., Quetier F., Saurin W., Scarpelli C.,
RA   Wincker P., Lander E.S., Weissenbach J., Roest Crollius H.;
RT   "Genome duplication in the teleost fish Tetraodon nigroviridis reveals
RT   the early vertebrate proto-karyotype.";
RL   Nature 431:946-957(2004).
RN   [2]
RP   NUCLEOTIDE SEQUENCE.
RG   Genoscope; Whitehead Institute Centre for Genome Research;
RL   Submitted (FEB-2004) to the EMBL/GenBank/DDBJ databases.
CC   -!- CAUTION: The sequence shown here is derived from an
CC       EMBL/GenBank/DDBJ whole genome shotgun (WGS) entry which is
CC       preliminary data.
DR   EMBL; CAAE01015001; CAG08859.1; -; Genomic_DNA.
FT   NON_TER       1       1
FT   NON_TER     600     600
SQ   SEQUENCE   600 AA;  66291 MW;  B2F28B2F3B27B1C4 CRC64;

  Query Match              5.3%;  Score 34;  DB 2;  Length 600;
  Best Local Similarity   100.0%;  Pred. No. 6.6e-24;
  Matches   34;  Conservative    0;  Mismatches    0;  Indels    0;  Gaps    0;

Qy        501 ELRKLNKVISDLTPVSELPLTARPRSRKEKNKLA 534
              ||||||||||||||||||||||||||||||||||
Db        421 ELRKLNKVISDLTPVSELPLTARPRSRKEKNKLA 454


RESULT 11
Q6UAY4_WSSV
ID   Q6UAY4_WSSV PRELIMINARY;      PRT;   167 AA.
AC   Q6UAY4;
DT   05-JUL-2004 (TrEMBLrel. 27, Created)
DT   05-JUL-2004 (TrEMBLrel. 27, Last sequence update)
DT   05-JUL-2004 (TrEMBLrel. 27, Last annotation update)
DE   Wsv285 (Fragment).
OS   White spot syndrome virus (WSSV).
OC   Viruses; dsDNA viruses, no RNA stage; Nimaviridae; Whispovirus.
OX   NCBI_TaxID=92652;
RN   [1]
RP   NUCLEOTIDE SEQUENCE.
RA   Rajendran K.V., Mukherjee S.C., Vijayan K.K., Jung S.J., Kim Y.J.,
RA   Oh M.J.;
RL   Submitted (AUG-2003) to the EMBL/GenBank/DDBJ databases.
DR   EMBL; AY374443; AAQ92041.1; -; Genomic_DNA.
FT   NON_TER       1       1
FT   NON_TER     167     167
SQ   SEQUENCE   167 AA;  18918 MW;  BFCB713D3B1D081A CRC64;

  Query Match              1.7%;  Score 11;  DB 2;  Length 167;
  Best Local Similarity   100.0%;  Pred. No. 0.093;
```

```
  Matches   11;  Conservative    0;  Mismatches    0;  Indels    0;  Gaps    0;

Qy        380 SENEEEEEEEE 390
              ||||||||||
Db         92 SENEEEEEEEE 102


RESULT 12
Q9QTF1_9BACU
ID   Q9QTF1_9BACU PRELIMINARY;      PRT;   215 AA.
AC   Q9QTF1;
DT   01-MAY-2000 (TrEMBLrel. 13, Created)
DT   01-MAY-2000 (TrEMBLrel. 13, Last sequence update)
DT   01-OCT-2002 (TrEMBLrel. 22, Last annotation update)
DE   HHNBV-XIA.
GN   Name=HHNBV-XIA;
OS   unidentified baculovirus.
OC   Viruses; dsDNA viruses, no RNA stage; Baculoviridae.
OX   NCBI_TaxID=10469;
RN   [1]
RP   NUCLEOTIDE SEQUENCE.
RA   Xia C., Liu J.;
RL   Submitted (DEC-1998) to the EMBL/GenBank/DDBJ databases.
DR   EMBL; AB021155; BAA83091.1; -; Genomic_DNA.
SQ   SEQUENCE   215 AA;  24441 MW;  913D2D116BBB7B5B CRC64;

  Query Match             1.7%;  Score 11;  DB 2;  Length 215;
  Best Local Similarity   100.0%;  Pred. No. 0.12;
  Matches   11;  Conservative    0;  Mismatches    0;  Indels    0;  Gaps    0;

Qy        380 SENEEEEEEEE 390
              ||||||||||
Db        115 SENEEEEEEEE 125


RESULT 13
Q91B56_WSSV
ID   Q91B56_WSSV PRELIMINARY;      PRT;   358 AA.
AC   Q91B56;
DT   01-DEC-2001 (TrEMBLrel. 19, Created)
DT   01-DEC-2001 (TrEMBLrel. 19, Last sequence update)
DT   01-OCT-2003 (TrEMBLrel. 25, Last annotation update)
DE   Hypothetical protein.
OS   White spot syndrome virus (WSSV).
OC   Viruses; dsDNA viruses, no RNA stage; Nimaviridae; Whispovirus.
OX   NCBI_TaxID=92652;
RN   [1]
RP   NUCLEOTIDE SEQUENCE.
RA   Shi Z., Bonami J.-R.;
RL   Submitted (JAN-2001) to the EMBL/GenBank/DDBJ databases.
DR   EMBL; AF343568; AAL24459.1; -; Genomic_DNA.
KW   Hypothetical protein.
SQ   SEQUENCE   358 AA;  40861 MW;  FFF05A8A812B2680 CRC64;

  Query Match             1.7%;  Score 11;  DB 2;  Length 358;
  Best Local Similarity   100.0%;  Pred. No. 0.18;
  Matches   11;  Conservative    0;  Mismatches    0;  Indels    0;  Gaps    0;

Qy        380 SENEEEEEEEE 390
              ||||||||||
Db        258 SENEEEEEEEE 268


RESULT 14
Q6BKF1_DEBHA
ID   Q6BKF1_DEBHA PRELIMINARY;      PRT;   468 AA.
AC   Q6BKF1;
DT   25-OCT-2004 (TrEMBLrel. 28, Created)
DT   25-OCT-2004 (TrEMBLrel. 28, Last sequence update)
DT   25-OCT-2004 (TrEMBLrel. 28, Last annotation update)
DE   Similar to CA5208|IPF2361 Candida albicans IPF2361.
GN   OrderedLocusNames=DEHA0F23914g;
OS   Debaryomyces hansenii (Yeast) (Torulaspora hansenii).
OC   Eukaryota; Fungi; Ascomycota; Saccharomycotina; Saccharomycetes;
OC   Saccharomycetales; Saccharomycetaceae; Debaryomyces.
OX   NCBI_TaxID=4959;
RN   [1]
RP   NUCLEOTIDE SEQUENCE [LARGE SCALE GENOMIC DNA].
RC   STRAIN=ATCC 36239 / CBS 767;
RX   PubMed=15229592; DOI=10.1038/nature02579;
RA   Dujon B., Sherman D., Fischer G., Durrens P., Casaregola S.,
RA   Lafontaine I., de Montigny J., Marck C., Neuveglise C., Talla E.,
RA   Goffard N., Frangeul L., Aigle M., Anthouard V., Babour A., Barbe V.,
RA   Barnay S., Blanchin S., Beckerich J.-M., Beyne E., Bleykasten C.,
RA   Boisrame A., Boyer J., Cattolico L., Confanioleri F., de Daruvar A.,
RA   Despons L., Fabre E., Fairhead C., Ferry-Dumazet H., Groppi A.,
RA   Hantraye F., Hennequin C., Jauniaux N., Joyet P., Kachouri R.,
RA   Kerrest A., Koszul R., Lemaire M., Lesur I., Ma L., Muller H.,
RA   Nicaud J.-M., Nikolski M., Oztas S., Ozier-Kalogeropoulos O.,
RA   Pellenz S., Potier S., Richard G.-F., Straub M.-L., Suleau A.,
RA   Swennen D., Tekaia F., Wesolowski-Louvel M., Westhof E., Wirth B.,
RA   Zeniou-Meyer M., Zivanovic Y., Bolotin-Fukuhara M., Thierry A.,
RA   Bouchier C., Caudron B., Scarpelli C., Gaillardin C., Weissenbach J.,
RA   Wincker P., Souciet J.-L.;
RT   "Genome evolution in yeasts.";
RL   Nature 430:35-44(2004).
DR   EMBL; CR382138; CAG89724.1; -; Genomic_DNA.
DR   InterPro; IPR000313; PWWP.
DR   Pfam; PF00855; PWWP; 1.
DR   SMART; SM00293; PWWP; 1.
DR   PROSITE; PS50812; PWWP; 1.
KW   Complete proteome.
SQ   SEQUENCE   468 AA;  53829 MW;  F1C367DFC5FE8117 CRC64;

  Query Match             1.7%;  Score 11;  DB 2;  Length 468;
  Best Local Similarity   100.0%;  Pred. No. 0.23;
  Matches   11;  Conservative    0;  Mismatches    0;  Indels    0;  Gaps    0;

Qy        383 EEEEEEEEDYE 393
              ||||||||||
Db        190 EEEEEEEEDYE 200


RESULT 15
O17941_CAEEL
ID   O17941_CAEEL PRELIMINARY;      PRT;   517 AA.
AC   O17941; O18079;
DT   01-JAN-1998 (TrEMBLrel. 05, Created)
DT   01-NOV-1998 (TrEMBLrel. 08, Last sequence update)
DT   13-SEP-2005 (TrEMBLrel. 31, Last annotation update)
DE   Hypothetical protein K08F9.4.
GN   ORFNames=K08F9.4;
OS   Caenorhabditis elegans.
OC   Eukaryota; Metazoa; Nematoda; Chromadorea; Rhabditida; Rhabditoidea;
OC   Rhabditidae; Peloderinae; Caenorhabditis.
OX   NCBI_TaxID=6239;
RN   [1]
RP   NUCLEOTIDE SEQUENCE.
RC   STRAIN=Bristol N2;
RA   Mortimore B.J.;
RL   Submitted (DEC-1996) to the EMBL/GenBank/DDBJ databases.
RN   [2]
RP   NUCLEOTIDE SEQUENCE [LARGE SCALE GENOMIC DNA].
RC   STRAIN=Bristol N2;
RX   MEDLINE=99069613; PubMed=9851916; DOI=10.1126/science.282.5396.2012;
RG   The C. elegans sequencing consortium;
RT   "Genome sequence of the nematode C. elegans: a platform for
RT   investigating biology.";
RL   Science 282:2012-2018(1998).
DR   EMBL; Z83239; CAB05811.1; -; Genomic_DNA.
DR   EMBL; Z81099; CAB03189.1; -; Genomic_DNA.
DR   EMBL; Z81099; CAB05811.1; JOINED; Genomic_DNA.
DR   EMBL; Z83239; CAB03189.1; JOINED; Genomic_DNA.
DR   PIR; T23499; T23499.
DR   Ensembl; K08F9.4; Caenorhabditis elegans.
DR   WormBase; WBGene00010687; K08F9.4.
```

DR WormPep; K08F9.4; CE11948.
KW Complete proteome; Hypothetical protein.
SQ SEQUENCE 517 AA; 59610 MW; 68E9E8909D5808FD CRC64;

Query Match 1.7%; Score 11; DB 2; Length 517;
Best Local Similarity 100.0%; Pred. No. 0.25;
Matches 11; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

```
Qy        380 SENEEEEEEEE 390
              ||||||||||| 
Db         21 SENEEEEEEEE 31
```

Search completed: February 28, 2006, 09:09:46
Job time : 233 secs

GenCore version 5.1.7

OM protein - protein search, using sw model

Run on: February 28, 2006, 08:37:25 ; Search time 153.45 Seconds
(without alignments)
1282.772 Million cell updates/sec

Title: US-10-717-665A-44_COPY_1_448
Perfect score: 2364
Sequence: 1 MPQPSVSGMDPPFGDAFRSH.....GKRRYFWEYSEQLTPSQQER 448

Scoring table: BLOSUM62
Gapop 10.0 , Gapext 0.5

Searched: 2443163 seqs, 439378781 residues

Total number of hits satisfying chosen parameters: 2443163

Minimum DB seq length: 0
Maximum DB seq length: 2000000000

Post-processing: Minimum Match 0%
Maximum Match 100%
Listing first 45 summaries

Database : A_Geneseq_21:*
1: geneseqp1980s:*
2: geneseqp1990s:*
3: geneseqp2000s:*
4: geneseqp2001s:*
5: geneseqp2002s:*
6: geneseqp2003as:*
7: geneseqp2003bs:*
8: geneseqp2004s:*
9: geneseqp2005s:*

Pred. No. is the number of results predicted by chance to have a score greater than or equal to the score of the result being printed, and is derived by analysis of the total score distribution.

SUMMARIES

Result No.	Score	% Query Match	Length	DB	ID	Description
1	2364	100.0	639	7	ADK65805	Adk65805 Angiogene
2	2148	90.9	417	8	ADQ66472	Adq66472 Novel hum
3	1268	53.6	256	5	ADK35935	Adk35935 Novel hum
4	865	36.6	168	5	ABP64928	Abp64928 Human pro
5	557	23.6	296	8	ADO20373	Ado20373 Human PRO
6	190	8.0	1162	3	AAY96255	Aay96255 Kaposi's
7	190	8.0	1162	3	AAY58500	Aay58500 HHV8 ORF
8	190	8.0	1162	4	AAB62331	Aab62331 Amino aci
9	190	8.0	1162	5	ABB05621	Abb05621 Kaposi's
10	190	8.0	1162	8	ADJ65096	Adj65096 HHV8 late
11	190	8.0	1162	9	ADV68154	Adv68154 Kaposi's
12	185.5	7.8	712	7	ADE61623	Ade61623 Rat Prote
13	185.5	7.8	712	7	ADD46090	Add46090 Rat Prote
14	185.5	7.8	712	7	ADE57828	Ade57828 Rat Prote
15	185.5	7.8	712	7	ADE57830	Ade57830 Rat Prote
16	179	7.6	842	4	AAB85725	Aab85725 Rat TBP-b
17	169.5	7.2	735	6	AAE37016	Aae37016 Human nuc
18	168	7.1	712	2	AAW30749	Aaw30749 Rat YT521
19	168	7.1	712	7	ADE56302	Ade56302 Rat Prote
20	165	7.0	520	8	ADP46649	Adp46649 Human col
21	165	7.0	568	8	ADP46648	Adp46648 Human col
22	165	7.0	706	7	ADE61625	Ade61625 Human Pro
23	165	7.0	706	7	ADD46092	Add46092 Human Pro
24	165	7.0	707	2	AAR79912	Aar79912 Human nuc
25	165	7.0	707	2	AAW84052	Aaw84052 Human V3
26	165	7.0	707	4	AAB48964	Aab48964 Human nuc
27	165	7.0	707	7	ADD49220	Add49220 Human nuc
28	165	7.0	707	8	ADJ58974	Adj58974 Human nuc
29	165	7.0	707	8	ADP54086	Adp54086 Human PRO
30	165	7.0	707	9	ADY19864	Ady19864 PRO polyp
31	165	7.0	710	8	ABM80398	Abm80398 Tumour-as
32	163	6.9	354	9	ADZ72253	Adz72253 Plasmodiu
33	162.5	6.9	1471	8	ADP25445	Adp25445 Plasmodiu
34	162	6.9	167	4	ABG11270	Abg11270 Novel hum
35	161	6.8	831	3	AAY58002	Aay58002 Saccharom
36	160.5	6.8	275	8	ADJ58999	Adj58999 Human nuc
37	160.5	6.8	382	4	ABG04351	Abg04351 Novel hum
38	160.5	6.8	1974	8	ADN22802	Adn22802 Bacterial
39	160	6.8	927	5	AAE18907	Aae18907 Human PAS
40	160	6.8	1972	2	AAW81171	Aaw81171 Human BAZ
41	160	6.8	1972	6	ABR64241	Abr64241 Angiogene
42	160	6.8	1972	8	ADP54420	Adp54420 Human PRO
43	159.5	6.7	3147	9	AEB22180	Aeb22180 Codon opt
44	159.5	6.7	3553	9	AEB22174	Aeb22174 Plasmodiu
45	159	6.7	1038	4	ABB70151	Abb70151 Drosophil

ALIGNMENTS

RESULT 1
ADK65805
ID ADK65805 standard; protein; 639 AA.
XX
AC ADK65805;
XX
DT 06-MAY-2004 (first entry)
XX
DE Angiogenesis-differentially expressed protein ANH0757.
XX
KW cytostatic; cardiant; vasotropic; antiarteriosclerotic;
KW angiogenesis inhibitor; angiogenesis stimulator; angiogenic index;
KW gene expression; cancer; coronary artery disease; myocardial ischemia;
KW coronary arteriosclerosis; forensic medicine.
XX
OS Homo sapiens.
XX
PN WO2003066831-A2.
XX
PD 14-AUG-2003.
XX
PF 07-FEB-2003; 2003WO-US003848.
XX
PR 07-FEB-2002; 2002US-00067482.
PR 10-JUN-2002; 2002US-00164595.
PR 16-AUG-2002; 2002US-0403649P.
PR 03-JAN-2003; 2003US-0437746P.
XX
PA (ORIG-) ORIGENE TECHNOLOGIES INC.
XX
PI Sun Z, Li X, Kovacs KF, Fan W, Jay G;
XX
DR WPI; 2003-731502/69.
DR N-PSDB; ADK65804.
XX
PT Determining the angiogenic index of a tissue or cell sample using
PT expression levels of differentially expressed genes, useful for
PT diagnosing or treating cancer, coronary artery disease, myocardial
PT ischemia and/or arteriosclerosis.
XX
PS Claim 23; SEQ ID NO 44; 296pp; English.
XX
CC The invention relates to a method of determining the angiogenic index of
CC a tissue or cell sample comprising assessing, in a sample, the expression
CC levels of one or more differentially-expressed gene from any of 34 DNA
CC sequences, given in the specification, where the levels are indicative of
CC the angiogenic index. The methods and compositions of the present

```
CC   invention are useful for diagnosing, preventing and/or treating cancer,
CC   coronary artery disease, myocardial ischemia or coronary
CC   arteriosclerosis. They can also be used in research, drug discovery and
CC   forensic medicine involving angiogenesis. This sequence corresponds to
CC   one of the differentially expressed proteins of the invention.
XX
SQ   Sequence 639 AA;

  Query Match             100.0%;  Score 2364;  DB 7;  Length 639;
  Best Local Similarity   100.0%;  Pred. No. 1.8e-188;
  Matches  448;  Conservative    0;  Mismatches    0;  Indels    0;  Gaps    0;

Qy          1 MPQPSVSGMDPPFGDAFRSHTFSEQTLMSTDLLANSSDPDFMYELDREMNYQQNPRDNFL 60
              ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db          1 MPQPSVSGMDPPFGDAFRSHTFSEQTLMSTDLLANSSDPDFMYELDREMNYQQNPRDNFL 60

Qy         61 SLEDCKDIENLESFTDVLDNEGALTSNWEQWDTYCEDLTKYTKLTSCDIWGTKEVDYLGL 120
              ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db         61 SLEDCKDIENLESFTDVLDNEGALTSNWEQWDTYCEDLTKYTKLTSCDIWGTKEVDYLGL 120

Qy        121 DDFSSPYQDEEVISKTPTLAQLNSEDSQSVSDSLYYPDSLFSVKQNPLPSSFPGKKITSR 180
              ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db        121 DDFSSPYQDEEVISKTPTLAQLNSEDSQSVSDSLYYPDSLFSVKQNPLPSSFPGKKITSR 180

Qy        181 AAAPVCSSKTLQAEVPLSDCVQKASKPPSSTQIMVKTNMYHNEKVNFHVECKDYVKKAKV 240
              ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db        181 AAAPVCSSKTLQAEVPLSDCVQKASKPPSSTQIMVKTNMYHNEKVNFHVECKDYVKKAKV 240

Qy        241 KINPVQQSRPLLSQIHTDAAKENTCYCGAVAKRQEKKGMEPLQGHATPALPFKETQELLL 300
              ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db        241 KINPVQQSRPLLSQIHTDAAKENTCYCGAVAKRQEKKGMEPLQGHATPALPFKETQELLL 300

Qy        301 SPLPQEGPGSLAAGESSSLSASTSVSDSSQKKEEHNYSLFVSDNLGEQPTKCSPEEDEED 360
              ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db        301 SPLPQEGPGSLAAGESSSLSASTSVSDSSQKKEEHNYSLFVSDNLGEQPTKCSPEEDEED 360

Qy        361 EEDVDDEDHDEGFGSEHELSENEEEEEEEEDYEDDKDDDISDTFSEPGYENDSVEDLKEV 420
              ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db        361 EEDVDDEDHDEGFGSEHELSENEEEEEEEEDYEDDKDDDISDTFSEPGYENDSVEDLKEV 420

Qy        421 TSISSRKRGKRRYFWEYSEQLTPSQQER 448
              ||||||||||||||||||||||||||||
Db        421 TSISSRKRGKRRYFWEYSEQLTPSQQER 448


RESULT 2
ADQ66472
ID   ADQ66472 standard; protein; 417 AA.
XX
AC   ADQ66472;
XX
DT   07-OCT-2004  (first entry)
XX
DE   Novel human protein sequence #1445.
XX
KW   osteopathic; neuroprotective; nootropic; antiparkinsonian; cytostatic;
KW   gene therapy; diagnostic marker; morbid state; osteoporosis;
KW   neurological disease; Alzheimer's disease; Parkinson's disease; dementia;
KW   cancer.
XX
OS   Homo sapiens.
XX
PN   EP1440981-A2.
XX
PD   28-JUL-2004.
XX
PF   21-JAN-2004; 2004EP-00001196.
XX
PR   21-JAN-2003; 2003JP-00102206.
PR   09-MAY-2003; 2003JP-00131392.
XX
PA   (REAS-) RES ASSOC BIOTECHNOLOGY.
XX
PI   Isogai T,  Sugiyama T,  Otsuki T,  Wakamatsu A,  Sato H,  Ishii S;
PI   Yamamoto J,  Isono Y,  Nagai K,  Irie R;
XX
DR   WPI; 2004-535376/52.
DR   N-PSDB; ADQ64284.
XX
PT   Novel 2495 cDNA, useful for treating osteoporosis, neurological diseases,
PT   Alzheimer's diseases, Parkinson's diseases, dementia and various cancers.
XX
PS   Claim 1; SEQ ID NO 3633; 2449pp; English.
XX
CC   The invention relates to 2495 novel polynucleotides (I) and their encoded
CC   polypeptides, sequences hybridizing to these nucleotides, sequences
CC   encoding partial polypeptides and sequences having 70% or 90% identity to
CC   the nucleotide and protein sequences. The nucleotides and polypeptides
CC   are useful as diagnostic markers or therapeutic target for the diseases
CC   or morbid states. They are also useful for treating osteoporosis,
CC   neurological diseases, Alzheimer's diseases, Parkinson's diseases,
CC   dementia and various cancers. This sequence corresponds to a protein
CC   sequence of the invention.
XX
SQ   Sequence 417 AA;

  Query Match             90.9%;  Score 2148;  DB 8;  Length 417;
  Best Local Similarity   99.8%;  Pred. No. 1.1e-170;
  Matches  407;  Conservative    0;  Mismatches    1;  Indels    0;  Gaps    0;

Qy          1 MPQPSVSGMDPPFGDAFRSHTFSEQTLMSTDLLANSSDPDFMYELDREMNYQQNPRDNFL 60
              ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db          1 MPQPSVSGMDPPFGDAFRSHTFSEQTLMSTDLLANSSDPDFMYELDREMNYQQNPRDNFL 60

Qy         61 SLEDCKDIENLESFTDVLDNEGALTSNWEQWDTYCEDLTKYTKLTSCDIWGTKEVDYLGL 120
              ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db         61 SLEDCKDIENLESFTDVLDNEGALTSNWEQWDTYCEDLTKYTKLTSCDIWGTKEVDYLGL 120

Qy        121 DDFSSPYQDEEVISKTPTLAQLNSEDSQSVSDSLYYPDSLFSVKQNPLPSSFPGKKITSR 180
              ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db        121 DDFSSPYQDEEVISKTPTLAQLNSEDSQSVSDSLYYPDSLFSVKQNPLPSSFPGKKITSR 180

Qy        181 AAAPVCSSKTLQAEVPLSDCVQKASKPPSSTQIMVKTNMYHNEKVNFHVECKDYVKKAKV 240
              |||||||||||||||||||||||| |||||||||||||||||||||||||||||||||||
Db        181 AAAPVCSSKTLQAEVPLSDCVQKASKPTSSTQIMVKTNMYHNEKVNFHVECKDYVKKAKV 240

Qy        241 KINPVQQSRPLLSQIHTDAAKENTCYCGAVAKRQEKKGMEPLQGHATPALPFKETQELLL 300
              ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db        241 KINPVQQSRPLLSQIHTDAAKENTCYCGAVAKRQEKKGMEPLQGHATPALPFKETQELLL 300

Qy        301 SPLPQEGPGSLAAGESSSLSASTSVSDSSQKKEEHNYSLFVSDNLGEQPTKCSPEEDEED 360
              ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db        301 SPLPQEGPGSLAAGESSSLSASTSVSDSSQKKEEHNYSLFVSDNLGEQPTKCSPEEDEED 360

Qy        361 EEDVDDEDHDEGFGSEHELSENEEEEEEEEDYEDDKDDDISDTFSEPG 408
              ||||||||||||||||||||||||||||||||||||||||||||||||
Db        361 EEDVDDEDHDEGFGSEHELSENEEEEEEEEDYEDDKDDDISDTFSEPG 408


RESULT 3
ADK35935
ID   ADK35935 standard; protein; 256 AA.
XX
AC   ADK35935;
XX
DT   06-MAY-2004  (first entry)
XX
DE   Novel human polypeptide SeqID8017.
XX
KW   antiarthritic; antiparkinsonian; neuroprotective; nootropic;
KW   immunosuppressive; cytostatic; antipsoriatic; antiinflammatory;
KW   antibacterial; antiviral; antifungal; antiparasitic; gene therapy;
KW   arthritis; Parkinson's; Alzheimer's; autoimmune disease; cancer;
KW   psoriasis; inflammatory bowel disease; infection; bacteria; virus;
KW   fungus; parasite; human.
XX
```

```
OS   Homo sapiens.
XX
FH   Key             Location/Qualifiers
FT   Misc-difference 1. .256
FT                   /label= OTHER
FT                   /note= "OTHER= All Xaa's in this sequence are unknown
FT                   amino acids or the site of a stop codon within the DNA
FT                   sequence"
XX
PN   WO200216439-A2.
XX
PD   28-FEB-2002.
XX
PF   05-MAR-2001; 2001WO-US004941.
XX
PR   07-MAR-2000; 2000US-00519705.
PR   19-MAY-2000; 2000US-00574454.
XX
PA   (HYSE-) HYSEQ INC.
XX
PI   Tang YT,  Liu C,  Drmanac RT;
XX
DR   WPI; 2002-280918/32.
XX
PT   Isolated polynucleotide encoding bone marrow derived polypeptides useful
PT   for treating, e.g., Parkinson's, Alzheimer's, cancer, arthritis, Crohn's
PT   disease, and inflammatory bowel disease.
XX
PS   Claim 20; SEQ ID NO 8017; 504pp; English.
XX
CC   This invention relates to a novel isolated polynucleotide comprising a
CC   nucleotide sequence selected from one of 1680 sequences, a mature protein
CC   coding portion of them, an active domain of them and their complementary
CC   sequences. The invention may be useful for the production of compounds
CC   with an antiarthritic, antiparkinsonian, neuroprotective, nootropic,
CC   immunosuppressive, cytostatic, antipsoriatic, antiinflammatory,
CC   antibacterial, antiviral, antifungal or antiparasitic activity. In
CC   addition, the disclosed sequences may be useful for gene therapy. The
CC   polypeptides or their antibodies are useful for treating many diseases
CC   such as arthritis, Parkinson's, Alzheimer's, autoimmune diseases, cancer,
CC   psoriasis, inflammatory bowel disease and infections caused by bacteria,
CC   viruses, fungi or parasites. The present sequence is that of a human
CC   polypeptide of the invention.
XX
SQ   Sequence 256 AA;

  Query Match             53.6%;  Score 1268;  DB 5;  Length 256;
  Best Local Similarity   98.8%;  Pred. No. 2.3e-97;
  Matches  243;  Conservative    0;  Mismatches    3;  Indels    0;  Gaps    0;

Qy        163 VKQNPLPSSFPGKKITSRAAAPVCSSKTLQAEVPLSDCVQKASKPPSSTQIMVKTNMYHN 222
              ||||||||||||||| | |||||||||||||||||||||| ||||||||||||||||
Db          5 VKQNPLPSSFPGKKITIRLAAPVCSSKTLQAEVPLSDCVQKASKPTSSTQIMVKTNMYHN 64

Qy        223 EKVNFHVECKDYVKKAKVKINPVQQSRPLLSQIHTDAAKENTCYCGAVAKRQEKKGMEPL 282
              ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db         65 EKVNFHVECKDYVKKAKVKINPVQQSRPLLSQIHTDAAKENTCYCGAVAKRQEKKGMEPL 124

Qy        283 QGHATPALPFKETQELLLSPLPQEGPGSLAAGESSSLSASTSVSDSSQKKEEHNYSLFVS 342
              ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db        125 QGHATPALPFKETQELLLSPLPQEGPGSLAAGESSSLSASTSVSDSSQKKEEHNYSLFVS 184

Qy        343 DNLGEQPTKCSPEEDEEDEEDVDDEDHDEGFGSEHELSENEEEEEEEEDYEDDKDDDISD 402
              ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db        185 DNLGEQPTKCSPEEDEEDEEDVDDEDHDEGFGSEHELSENEEEEEEEEDYEDDKDDDISD 244

Qy        403 TFSEPG 408
              ||||||
Db        245 TFSEPG 250


RESULT 4
ABP64928
ID   ABP64928 standard; protein; 168 AA.
XX
AC   ABP64928;
XX
DT   25-FEB-2003  (first entry)
XX
DE   Human protein SEQ ID 588.
XX
KW   Human; expressed sequence tag; EST; haematopoietic disorder;
KW   central nervous system disease; viral infection;
KW   peripheral nervous system disease; non-healing wound; infectious disease;
KW   immune deficiency; immune disorder; bacterial infection; allergy; cancer;
KW   fungal infection; autoimmune disorder; coagulation disorder; nootropic;
KW   antiallergic; antiinflammatory; immunosuppressive; neuroprotective;
KW   cytostatic; haemostatic; virucide; antibacterial; fungicide;
KW   immunostimulant; cerebroprotective.
XX
OS   Homo sapiens.
XX
PN   WO200259260-A2.
XX
PD   01-AUG-2002.
XX
PF   16-NOV-2001; 2001WO-US042950.
XX
PR   17-NOV-2000; 2000US-00714936.
XX
PA   (HYSE-) HYSEQ INC.
XX
PI   Tang YT,  Goodrich RW,  Liu C,  Zhou P,  Asundi V,  Zhang J,  Zhao QA;
PI   Ren F,  Xue AJ,  Yang Y,  Wehrman T,  Drmanac RT;
XX
DR   WPI; 2002-590824/63.
DR   N-PSDB; ABQ99514.
XX
PT   New isolated polynucleotide, useful in research, diagnostic or
PT   therapeutic methods, e.g. preventing or treating disorders involving
PT   aberrant protein expression or biological activity.
XX
PS   Claim 20; SEQ ID NO 588; 394pp; English.
XX
CC   The present invention relates to novel human coding sequences (ABQ99268-
CC   ABQ99608) and proteins (ABP64682-ABP65022). The sequences are useful in
CC   therapeutic, diagnostic and research methods. The polynucleotides may be
CC   used in the field of molecular biology as hybridisation probes, primers
CC   for PCR, for chromosome and gene mapping, for the recombinant production
CC   of protein, or in generation of anti-sense DNA or RNA. The
CC   polynucleotides are useful in diagnostics as expressed sequence tags
CC   (ESTs) for identifying expressed genes or for physical mapping of the
CC   human genome. The proteins may be used as molecular weight markers, or as
CC   nutritional sources or supplements. The proteins may be used to maintain
CC   and expand cell population in a totipotential or pluripotential state
CC   useful for re-engineering damaged or diseased tissues, transplantation,
CC   manufacture of bio-pharmaceuticals or the development of bio-sensors. The
CC   polynucleotides and proteins are useful for preventing, treating or
CC   ameliorating disorders involving aberrant protein expression or
CC   biological activity, e.g. haematopoietic disorders, central/peripheral
CC   nervous system diseases, mechanical and traumatic disorders, non-healing
CC   wounds, immune deficiencies and disorders, infectious diseases caused by
CC   viral, bacterial or fungal infection, autoimmune disorders, allergic
CC   reactions and conditions, coagulation disorders, or cancer. The
CC   polynucleotide sequences of the invention were assembled from ESTs
CC   isolated mainly by sequencing by hybridisation, and in some cases,
CC   sequences obtained from one or more public databases. Note: The sequence
CC   data for this patent did not form part of the printed specification, but
CC   was obtained in electronic format directly from WIPO at
CC   ftp.wipo.int/pub/published_pct_sequences
XX
SQ   Sequence 168 AA;

  Query Match             36.6%;  Score 865;  DB 5;  Length 168;
  Best Local Similarity   100.0%;  Pred. No. 6.4e-64;
  Matches  164;  Conservative    0;  Mismatches    0;  Indels    0;  Gaps    0;
```

```
Qy        214 MVKTNMYHNEKVNFHVECKDYVKKAKVKINPVQQSRPLLSQIHTDAAKENTCYCGAVAKR 273
              ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db          1 MVKTNMYHNEKVNFHVECKDYVKKAKVKINPVQQSRPLLSQIHTDAAKENTCYCGAVAKR 60

Qy        274 QEKKGMEPLQGHATPALPFKETQELLLSPLPQEGPGSLAAGESSSLSASTSVSDSSQKKE 333
              ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db         61 QEKKGMEPLQGHATPALPFKETQELLLSPLPQEGPGSLAAGESSSLSASTSVSDSSQKKE 120

Qy        334 EHNYSLFVSDNLGEQPTKCSPEEDEEDEEDVDDEDHDEGFGSEH 377
              |||||||||||||||||||||||||||||||||||||||||||||
Db        121 EHNYSLFVSDNLGEQPTKCSPEEDEEDEEDVDDEDHDEGFGSEH 164
```

RESULT 5
ADO20373
ID ADO20373 standard; protein; 296 AA.
XX
AC ADO20373;
XX
DT 12-AUG-2004 (first entry)
XX
DE Human PRO polypeptide #633.
XX
KW Human; PRO; immune related disorder; systemic lupus erythematosus;
KW rheumatoid arthritis; osteoarthritis; juvenile chronic arthritis;
KW systemic sclerosis; Sjogren's syndrome; vasculitis; sarcoidosis;
KW autoimmune haemolytic anaemia; autoimmune thrombocytopenia; thyroiditis;
KW diabetes mellitus; renal disease; demyelinating disease;
KW central nervous system; peripheral nervous system;
KW demyelinating polyneuropathy; Guillain-Barre syndrome;
KW chronic inflammatory demyelinating polyneuropathy.
XX
OS Homo sapiens.
XX
PN WO2004043361-A2.
XX
PD 27-MAY-2004.
XX
PF 06-NOV-2003; 2003WO-US035268.
XX
PR 08-NOV-2002; 2002US-0425235P.
XX
PA (GETH) GENENTECH INC.
XX
PI Fong S, Dennis K, Clark H, Chiu H, Schoenfeld J, Williams PM;
PI Wood WI, Wu TD;
XX
DR WPI; 2004-420067/39.
DR N-PSDB; ADO20372.
XX
PT Novel PRO polypeptide e.g., PRO69614, PRO71106, or PRO86388 useful for
PT treating an immune related disorder such as systemic lupus erythematosus,
PT rheumatoid arthritis, osteoarthritis, juvenile chronic arthritis or
PT spondyloarthropathy.
XX
PS Claim 7; SEQ ID NO 1266; 1731pp; English.
XX
CC The invention relates to human PRO polypeptides and the polynucleotides
CC encoding them. The polypeptides and polynucleotides are useful for
CC treating and diagnosing immune related disorders in mammals. The immune
CC related disorders include systemic lupus erythematosus, rheumatoid
CC arthritis, osteoarthritis, juvenile chronic arthritis, systemic
CC sclerosis, Sjogren's syndrome, vasculitis, sarcoidosis, autoimmune
CC haemolytic anaemia, autoimmune thrombocytopenia, thyroiditis, diabetes
CC mellitus, immune-mediated renal disease, demyelinating diseases of the
CC central or peripheral nervous system, demyelinating polyneuropathy,
CC Guillain-Barre syndrome and chronic inflammatory demyelinating
CC polyneuropathy. This sequence represents a human PRO polypeptide of the
CC invention.
XX
SQ Sequence 296 AA;

```
  Query Match           23.6%;  Score 557;  DB 8;  Length 296;
  Best Local Similarity  99.0%;  Pred. No. 8.7e-38;
  Matches  104;  Conservative    0;  Mismatches    1;  Indels    0;  Gaps    0;

Qy        344 NLGEQPTKCSPEEDEEDEEDVDDEDHDEGFGSEHELSENEEEEEEEEDYEDDKDDDISDT 403
              |||||||||||||||||||||||||||||||||||||||||||| ||||||||||||||||
Db          1 NLGEQPTKCSPEEDEEDEEDVDDEDHDEGFGSEHELSENEEEEEVEEDYEDDKDDDISDT 60

Qy        404 FSEPGYENDSVEDLKEVTSISSRKRGKRRYFWEYSEQLTPSQQER 448
              |||||||||||||||||||||||||||||||||||||||||||||
Db         61 FSEPGYENDSVEDLKEVTSISSRKRGKRRYFWEYSEQLTPSQQER 105
```

RESULT 6
AAY96255
ID AAY96255 standard; protein; 1162 AA.
XX
AC AAY96255;
XX
DT 12-SEP-2003 (revised)
DT 11-SEP-2000 (first entry)
XX
DE Kaposi's sarcoma-associated herpesvirus LANA.
XX
KW Kaposi's sarcoma-associated herpesvirus; KSHV; rhadino virus;
KW latency-associated nuclear antigen; LANA; gamma-2 herpes virus;
KW Human herpes virus 8; HHV8; rhadino virus cis-acting element; RVCAE;
KW Kaposi's sarcoma; primary effusion lymphoma; PEL;
KW human immunodeficiency virus; HIV; multicentric Castleman's disease.
XX
OS Human herpesvirus 8.
XX
FH Key   Location/Qualifiers
FT Domain   14. .17
FT   /note= "nuclear localisation signal, NLS"
FT Domain   64. .70
FT   /note= "nuclear localisation signal, NLS"
FT Region   320. .429
FT   /note= "acidic repeat region"
FT Region   430. .549
FT   /note= "Gln, Glu, Pro-rich region"
FT Region   550. .589
FT   /note= "Gln, Glu, Pro, Arg-rich region"
FT Region   590. .759
FT   /note= "Gln, Glu, Asp-rich region"
FT Region   760. .840
FT   /note= "Gln, Glu-rich region"
XX
PN WO200029626-A1.
XX
PD 25-MAY-2000.
XX
PF 19-NOV-1999; 99WO-US027508.
XX
PR 19-NOV-1998; 98US-00109422.
PR 21-APR-1999; 99US-00298568.
XX
PA (KIEF/) KIEFF E D.
PA (BALL/) BALLESTAS M E.
PA (KAYE/) KAYE K M.
XX
PI Kieff ED, Ballestas ME, Kaye KM;
XX
DR WPI; 2000-387829/33.
DR N-PSDB; AAA30290.
XX
PT Treating or preventing a disease associated with rhodino virus infection
PT in a mammal which includes Kaposi's Sarcoma and Primary Effusion
PT Lymphoma.
XX
PS Disclosure; Fig 7; 70pp; English.
XX
CC The present sequence is the Kaposi's sarcoma-associated herpesvirus,

```
CC   (KSHV) latency-associated nuclear antigen (LANA). KSHV is also known as
CC   Human Herpes Virus 8 (HHV8) and belongs to the rhadino virus, or gamma-2
CC   herpes virus class. The LANA protein is necessary for the efficient
CC   persistence of rhadino virus DNA in mammalian cells. Persistent rhadino
CC   virus infection is implicated in a variety of diseases e.g. Kaposi's
CC   Sarcoma (KS), Primary Effusion Lymphoma (PEL) and multicentric
CC   Castleman's disease. In addition, KS is a common malignancy in HIV
CC   patients. KSHV persists in host cells in a latent form. One of the few
CC   genes expressed from the latent viral DNA is LANA. LANA associates with
CC   both human chromosomes and with the rhadino virus cis-acting element
CC   (RVCAE), thereby providing a tethering function: the KSHV DNA episome is
CC   "tied" to the host chromosomes. This allows the viral DNA to persist in
CC   the host cell. The present sequence may be used to screen and identify
CC   molecules that inhibit LANA interaction with RVCAE, thereby interfering
CC   with the latency cycle of this virus. Potential antiviral treatments for
CC   the above mentioned diseases may therefore be based on LANA deregulation.
CC   (Updated on 12-SEP-2003 to standardise OS field)
XX
SQ   Sequence 1162 AA;

  Query Match             8.0%;  Score 190;  DB 3;  Length 1162;
  Best Local Similarity   24.2%;  Pred. No. 3.4e-06;
  Matches   94;  Conservative   50;  Mismatches  137;  Indels  108;  Gaps   14;

Qy        124 SSPYQDEEVISKTPTLAQLNSEDSQSVSDSLYYPDSLFSVKQNPLPSSFPGKKITSRAAA 183
              |||      :|   |       |: | | :|    |  | :: || ||  |:  :|
Db        111 SSPIPPSHPVSPGTT-------DTHSPSPALPPTQSPESSQRPPL-SSPTGRPDSSTPMR 162

Qy        184 PVCSSKTLQAEVPLSDCVQKASK------PPSSTQIMVKTNMYHNEKVNFHVECKDYVKK 237
              | | :|     | :   :  ||       ||: : : |  :                 :
Db        163 PPPSQQTTPPHSPTTPPPEPPSKSSPDSLAPSTLRSLRKRRLSS--------------PQ 208

Qy        238 AKVKINPVQQSRPLL--------SQIHTDAAKENTCYCGAVAKRQEKKGMEPLQ----GH 285
                :||: || |:          ::   | |:  | |: :       |      |
Db        209 GPSTLNPICQSPPVSPPRCDFANRSVYPPWATESPIYVGSSSDGDTPPRQPPTSPISIGS 268

Qy        286 ATPALPF--KETQELLL--------SPLPQEGPGSLAAGESSSLSASTSVSDSSQKKEEH 335
              ::|:      :|  |:|        |   :|   :  |||    :    :|  ||  ::
Db        269 SSPSEGSWGDDTAMLVLLAEIAEEASKNEKECSENNQAGED---NGDNEISKESQVDKDD 325

Qy        336 NYSLFVSDNLGEQPTKCSPEEDEEDEE--------------------------DVDDED 368
              |      ||  ::  : : ||||||:|                          : |||:
Db        326 N------DNKDDEEEQETDEEDEEDDEEDDEEDDEEDDEEDDEEDDEEDDEEDDEEDDEE 379

Qy        369 HDEGFGSEHELSENEEEEEEEEDYEDDKDDDISDTFSEPGYENDSVEDLKE--------- 419
               ||    | :  |:|||:|||||| ||| |:|  |       |:|  || ||
Db        380 DDEEDDEEEDEEEDEEEDEEEEDEDDDDEDNEDE------EDDEEEDKKEDEEDGGDGN 433

Qy        420 -VTSISSRKRGKRRYFWEYSEQLTPSQQE 447
               || | :: :        :|  | |||
Db        434 KTLSIQSSQQQQE------PQQQEPQQQE 456


RESULT 7
AAY58500
ID   AAY58500 standard; protein; 1162 AA.
XX
AC   AAY58500;
XX
DT   06-AUG-2003  (revised)
DT   10-APR-2000  (first entry)
XX
DE   HHV8 ORF 73 protein, SEQ ID NO:21.
XX
KW   HHV8; detection; diagnosis; Kaposi's sarcoma; AIDS; immunogen; antigen.
XX
OS   Human herpesvirus 8.
XX
FH   Key             Location/Qualifiers
FT   Misc-difference 96
FT                   /label= unknown
XX
PN   WO9961909-A2.
XX
PD   02-DEC-1999.
XX
PF   26-MAY-1999;   99WO-US011407.
XX
PR   26-MAY-1998;   98US-0086695P.
XX
PA   (USSH ) US DEPT HEALTH & HUMAN SERVICES.
XX
PI   Pau C;
XX
DR   WPI; 2000-097142/08.
XX
PT   New methods and compositions for the detection of human herpesvirus.
XX
PS   Claim 2; Page 59-62; 68pp; English.
XX
CC   Sequences AAY58480-Y58532 represent immunogenic polypeptides derived from
CC   human herpes virus type 8 (HHV8, a gammaherpesvirus). HHV8 plays an
CC   important role in the pathogenesis of AIDS-related Kaposi's sarcoma. The
CC   invention relates to a novel method of detecting the presence of human
CC   herpesvirus 8 in a biological sample using peptides representative of
CC   dominant antigenic regions of HHV8. The method comprises contacting one
CC   or more isolated, immunogenic HHV8 peptides with an antibody-containing
CC   biological sample, and detecting the formation of a complex between the
CC   peptide and the antibody. The presence of a peptide-antibody complex
CC   indicates the presence of human herpesvirus 8. The detection of HHV8
CC   infection can be used to diagnose AIDS-associated Kaposi's sarcoma. The
CC   HHV8-specific antibodies are useful therapeutically when for the passive
CC   immunisation of a human against HHV8 infection, thereby reducing HHV8
CC   related disease. The detection assays are highly specific, sensitive and
CC   accurate. Early detection and treatment of Kaposi's sarcoma could
CC   diminish the severity of symptoms related to AIDS and the sensitive
CC   techniques could reduce erroneous characterisations of skin disorders.
CC   Previous assays for HHV8 antibodies such as immunofluorescence assays,
CC   immunoblots and enzyme immunosassays lack the sensitivity and accuracy
CC   needed for reliable diagnosis of Kaposi's sarcoma. Further advantages of
CC   the assays are that reproducible results are obtained and the method is
CC   suitable for rapid throughput and screening of samples economically.
CC   (Updated on 06-AUG-2003 to correct OS field.)
XX
SQ   Sequence 1162 AA;

  Query Match             8.0%;  Score 190;  DB 3;  Length 1162;
  Best Local Similarity   24.2%;  Pred. No. 3.4e-06;
  Matches   94;  Conservative   50;  Mismatches  137;  Indels  108;  Gaps   14;

Qy        124 SSPYQDEEVISKTPTLAQLNSEDSQSVSDSLYYPDSLFSVKQNPLPSSFPGKKITSRAAA 183
              |||      :|   |       |: | | :|    |  | :: || ||  |:  :|
Db        111 SSPIPPSHPVSPGTT-------DTHSPSPALPPTQSPESSQRPPL-SSPTGRPDSSTPMR 162

Qy        184 PVCSSKTLQAEVPLSDCVQKASK------PPSSTQIMVKTNMYHNEKVNFHVECKDYVKK 237
              | | :|     | :   :  ||       ||: : : |  :                 :
Db        163 PPPSQQTTPPHSPTTPPPEPPSKSSPDSLAPSTLRSLRKRRLSS--------------PQ 208

Qy        238 AKVKINPVQQSRPLL--------SQIHTDAAKENTCYCGAVAKRQEKKGMEPLQ----GH 285
                :||: || |:          ::   | |:  | |: :       |      |
Db        209 GPSTLNPICQSPPVSPPRCDFANRSVYPPWATESPIYVGSSSDGDTPPRQPPTSPISIGS 268

Qy        286 ATPALPF--KETQELLL--------SPLPQEGPGSLAAGESSSLSASTSVSDSSQKKEEH 335
              ::|:      :|  |:|        |   :|   :  |||    :    :|  ||  ::
Db        269 SSPSEGSWGDDTAMLVLLAEIAEEASKNEKECSENNQAGED---NGDNEISKESQVDKDD 325

Qy        336 NYSLFVSDNLGEQPTKCSPEEDEEDEE--------------------------DVDDED 368
              |      ||  ::  : : ||||||:|                          : |||:
Db        326 N------DNKDDEEEQETDEEDEEDDEEDDEEDDEEDDEEDDEEDDEEDDEEDDEEDDEE 379

Qy        369 HDEGFGSEHELSENEEEEEEEEDYEDDKDDDISDTFSEPGYENDSVEDLKE--------- 419
               ||    | :  |:|||:|||||| ||| |:|  |       |:|  || ||
Db        380 DDEEDDEEEDEEEDEEEDEEEEDEDDDDEDNEDE------EDDEEEDKKEDEEDGGDGN 433

Qy        420 -VTSISSRKRGKRRYFWEYSEQLTPSQQE 447
               || | :: :        :|  | |||
```

```
Db        434 KTLSIQSSQQQQE------PQQQEPQQQE 456
```

RESULT 8
AAB62331
ID   AAB62331 standard; protein; 1162 AA.
XX
AC   AAB62331;
XX
DT   06-AUG-2003 (revised)
DT   29-JUN-2001 (first entry)
XX
DE   Amino acid sequence of KSHV tethering protein LANA.
XX
KW   Histone H1; tethering protein; LANA; gene therapy; multiple sclerosis;
KW   Parkinson's disease; Huntington disease; diabetes; human herpesvirus 8;
KW   KSHV; latency-associated nuclear antigen; LANA.
XX
OS   Human herpesvirus 8.
XX
PN   WO200125484-A2.
XX
PD   12-APR-2001.
XX
PF   29-SEP-2000; 2000WO-US026908.
XX
PR   01-OCT-1999;   99US-00410399.
XX
PA   (UNMI) UNIV MICHIGAN.
XX
PI   Robertson ES, Cotter MA;
XX
DR   WPI; 2001-281736/29.
DR   N-PSDB; AAF82901.
XX
PT   A composition for use in gene therapy comprises an expression vector that
PT   includes a nucleic acid sequence encoding a nucleic acid binding protein.
XX
PS   Disclosure; Fig 9B; 60pp; English.
XX
CC   The invention provides a composition comprising nucleic acid, histone H1
CC   protein and expression vector operationally encoding a protein suitable
CC   for tethering the nucleic acid to the histone H1 protein, where the
CC   tethering protein is LANA. The composition is useful in aiding the
CC   retention of the viral DNA in the host cell. The viral vector encodes a
CC   protein suitable for tethering DNA to Histone H1. Methods for screening
CC   for compounds which are agonistic or antagonistic for the tethering of
CC   viral proteins to histone H1 and DNA binding sites are useful for
CC   developing the method of viral transfer. The composition has applications
CC   to gene therapy, including the treatment of multiple sclerosis,
CC   Parkinson's disease, Huntington disease and diabetes. The present
CC   sequence represents the amino acid sequence of the Kaposi's sarcoma
CC   associated herpesvirus (human herpesvirus 8) latency-associated nuclear
CC   antigen (LANA), which acts as a tethering protein. (Updated on 06-AUG-
CC   2003 to correct OS field.)
XX
SQ   Sequence 1162 AA;

```
  Query Match             8.0%;  Score 190;  DB 4;  Length 1162;
  Best Local Similarity   24.2%;  Pred. No. 3.4e-06;
  Matches   94;  Conservative   50;  Mismatches  137;  Indels  108;  Gaps   14;

Qy        124 SSPYQDEEVISKTPTLAQLNSEDSQSVSDSLYYPDSLFSVKQNPLPSSFPGKKITSRAAA 183
              |||      :|   |          |: | | :|    |  | :: || ||  |:  :|
Db        111 SSPIPPSHPVSPGTT-------DTHSPSPALPPTQSPESSQRPPL-SSPTGRPDSSTPMR 162

Qy        184 PVCSSKTLQAEVPLSDCVQKASK------PPSSTQIMVKTNMYHNEKVNFHVECKDYVKK 237
              |  | :|     | :   :  ||       ||: : : |  :                :
Db        163 PPPSQQTTPPHSPTTPPPEPPSKSSPDSLAPSTLRSLRKRRLSS--------------PQ 208

Qy        238 AKVKINPVQQSRPLL--------SQIHTDAAKENTCYCGAVAKRQEKKGMEPLQ----GH 285
                :||: || |:          ::   | |:  | |: :          |      |
Db        209 GPSTLNPICQSPPVSPPRCDFANRSVYPPWATESPIYVGSSSDGDTPPRQPPTSPISIGS 268

Qy        286 ATPALPF--KETQELLL--------SPLPQEGPGSLAAGESSSLSASTSVSDSSQKKEEH 335
              ::|:      :|  |:|        |   :|   : |||    :    :|  ||  ::
Db        269 SSPSEGSWGDDTAMLVLLAEIAEEASKNEKECSENNQAGED---NGDNEISKESQVDKDD 325

Qy        336 NYSLFVSDNLGEQPTKCSPEEDEEDEE--------------------------DVDDED 368
              |      ||  ::  : : ||||||:|                          : |||:
Db        326 N------DNKDDEEEQETDEEDEEDDEEDDEEDDEEDDEEDDEEDDEEDDEEDDEE 379

Qy        369 HDEGFGSEHELSENEEEEEEEEDYEDDKDDDISDTFSEPGYENDSVEDLKE--------- 419
               ||    | :  |:|||:||||| ||| |:|  |       |:|  || ||
Db        380 DDEEDDEEEDEEEDEEEDEEEEDEEDDDDEDNEDE------EDDEEEDKKEDEEDGGDGN 433

Qy        420 -VTSISSRKRGKRRYFWEYSEQLTPSQQE 447
               || | :: :        :|  | |||
Db        434 KTLSIQSSQQQQE------PQQQEPQQQE 456

RESULT 9
ABB05621
ID   ABB05621 standard; protein; 1162 AA.
XX
AC   ABB05621;
XX
DT   25-APR-2002  (first entry)
XX
DE   Kaposi's sarcoma-associated herpesvirus LANA protein.
XX
KW   Kaposi's sarcoma-associated herpesvirus; KSHV; LANA; RVCAE; PEL;
KW   KSHV terminal repeat; rhadino virus cis acting element; episome;
KW   primary effusion lymphoma; latency-associated nuclear antigen;
KW   gene therapy; gene transfer.
XX
OS   Human herpesvirus 8.
XX
PN   US6322792-B1.
XX
PD   27-NOV-2001.
XX
PF   21-APR-1999;   99US-00298568.
XX
PR   19-NOV-1998;   98US-0109422P.
XX
PA   (KIEF/) KIEFF E D.
XX
PI   Kieff ED,  Ballestas ME,  Kaye KM;
XX
DR   WPI; 2002-153769/20.
DR   N-PSDB; ABA93487.
XX
PT   System for episomal retention of plasmids in mammalian cells, useful in
PT   gene therapy, comprises rhadinoviral LANA and RVCAE sequences.
XX
PS   Disclosure; Fig 7; 27pp; English.
XX
CC   The present invention describes a system (A) for maintaining a plasmid as
CC   an episome in mammalian cells, comprising the rhadinoviral sequence LANA
CC   (latency-associated nuclear antigen) of 3489 base pairs (see ABA93487,
CC   S1) expressed in the cell, and the rhadinoviral sequence RVCAE
CC   (rhadinoviral cis-acting element) of 801 base pairs (see ABA93488, S2)
CC   present in the plasmid. Also describes is a method for maintaining a
CC   closed circular DNA in a cell by expressing (S1) in the cells and having
CC   (S2) as a cis-acting and maintenance sequence in the DNA. (A) is
CC   particularly used in gene therapy (or other gene transfer applications)
CC   that uses mammalian cells in which LANA is expressed. (A) improves
CC   persistence of gene therapy vectors in cells. The present sequence
CC   represents Kaposi's sarcoma-associated herpesvirus (KSHV, also called
CC   human herpesvirus 8) LANA protein, which is used in the exemplification
CC   of the present invention
XX
SQ   Sequence 1162 AA;

```
  Query Match   8.0%; Score 190; DB 5; Length 1162;
```
```

```
  Best Local Similarity   24.2%;  Pred. No. 3.4e-06;
  Matches   94;  Conservative   50;  Mismatches  137;  Indels  108;  Gaps   14;

Qy        124 SSPYQDEEVISKTPTLAQLNSEDSQSVSDSLYYPDSLFSVKQNPLPSSFPGKKITSRAAA 183
              |||       :|    |        |: | | :|    |  | :: || ||  |:  :|
Db        111 SSPIPPSHPVSPGTT-------DTHSPSPALPPTQSPESSQRPPL-SSPTGRPDSSTPMR 162

Qy        184 PVCSSKTLQAEVPLSDCVQKASK------PPSSTQIMVKTNMYHNEKVNFHVECKDYVKK 237
              | | :|     | :   :  ||       ||: : : |  :             :
Db        163 PPPSQQTTPPHSPTTPPPEPPSKSSPDSLAPSTLRSLRKRRLSS-------------PQ 208

Qy        238 AKVKINPVQQSRPLL--------SQIHTDAAKENTCYCGAVAKRQEKKGMEPLQ----GH 285
                :||: || |:           ::   | |:  | |: :          |      |
Db        209 GPSTLNPICQSPPVSPPRCDFANRSVYPPWATESPIYVGSSSDGDTPPRQPPTSPISIGS 268

Qy        286 ATPALPF--KETQELLL--------SPLPQEGPGSLAAGESSSLSASTSVSDSSQKKEEH 335
              ::|:      :|  |:|        |   :|   :  |||    :    :|  ||  ::
Db        269 SSPSEGSWGDDTAMLVLLAEIAEEASKNEKECSENNQAGED---NGDNEISKESQVDKDD 325

Qy        336 NYSLFVSDNLGEQPTKCSPEEDEEDEE--------------------------DVDDED 368
              |      ||  ::  : : ||||||:|                            : |||:
Db        326 N------DNKDDEEEQETDEEDEEDDEEDDEEDDEEDDEEDDEEDDEEDDEEDDEEDDEE 379

Qy        369 HDEGFGSEHELSENEEEEEEEEDYEDDKDDDISDTFSEPGYENDSVEDLKE--------- 419
               ||    | :  |:|||:||||| ||| |:|  |       |:|  || ||
Db        380 DDEEDDEEDEEEDEEEDEEEEDEEDDDDEDNEDE------EDDEEEDKKEDEEDGGDGN 433

Qy        420 -VTSISSRKRGKRRYFWEYSEQLTPSQQE 447
               || | :: :        :|  | |||
Db        434 KTLSIQSSQQQQE------PQQQEPQQQE 456


RESULT 10
ADJ65096
ID   ADJ65096 standard; protein; 1162 AA.
XX
AC   ADJ65096;
XX
DT   20-MAY-2004  (first entry)
XX
DE   HHV8 latency-associated nuclear antigen, LANA.
XX
KW   HHV8; latency-associated nuclear antigen; LANA;
KW   primary effusion lymphoma; PEL; virucide; gene therapy; KSHV;
KW   Kaposi's sarcoma-associated herpesvirus; episome; RVCAE;
KW   rhadinovirus cis-acting element; rhadinovirus infection.
XX
OS   Human herpesvirus 8.
XX
PN   US2004037847-A1.
XX
PD   26-FEB-2004.
XX
PF   28-JUN-2001; 2001US-00894273.
XX
PR   19-NOV-1998;   98US-0109422P.
PR   21-APR-1999;   99US-00298568.
XX
PA   (KIEF/) KIEFF E D.
PA   (BALL/) BALLESTAS M E.
PA   (KAYE/) KAYE K M.
XX
PI   Kieff ED,  Ballestas ME,  Kaye KM;
XX
DR   WPI; 2004-191011/18.
DR   N-PSDB; ADJ65095.
XX
PT   Assays for compounds that modulate rhadino virus LANA action in trans on
PT   a unit of rhadino virus DNA to mediate efficient episome persistence,
PT   comprises detecting whether the compound inhibits expression of a gene
PT   resident on that episome.
XX
PS   Disclosure; SEQ ID NO 2; 28pp; English.
XX
CC   The invention relates to an assay for determining whether a compound
CC   modulates LANA (latency-associated nuclear antigen) from KSHV (Kaposi's
CC   sarcoma-associated herpesvirus, also known as HHV8, human herpesvirus 8)
CC   mediated persistence in a mammalian cell of a DNA episome having RVCAE
CC   (rhadinovirus cis-acting element) by detecting whether the compound
CC   inhibits expression of a gene resident on that episome. Also included are
CC   maintaining a closed circular DNA in a mammalian cell in which LANA is
CC   expressed (by using RVCAE as the cis-acting origin and maintenance
CC   sequence) and maintaining a plasmid as an episome in a mammalian cell
CC   (comprising: (a) Expressing LANA in the cell; and (b) having RVCAE
CC   resident on the plasmid). Compounds identified as interfering with the
CC   interaction between LANA and RVCAE are useful in treating or preventing
CC   rhadino virus infection (e.g. primary effusion lymphoma (PEL)). LANA acts
CC   in trans on a 0.8 Kb KSHV TR (terminal repeat) unit to mediate efficient
CC   episome persistence. The present sequence represents KSHV LANA.
XX
SQ   Sequence 1162 AA;

  Query Match             8.0%;  Score 190;  DB 8;  Length 1162;
  Best Local Similarity   24.2%;  Pred. No. 3.4e-06;
  Matches   94;  Conservative   50;  Mismatches  137;  Indels  108;  Gaps   14;

Qy        124 SSPYQDEEVISKTPTLAQLNSEDSQSVSDSLYYPDSLFSVKQNPLPSSFPGKKITSRAAA 183
              |||       :|    |        |: | | :|    |  | :: || ||  |:  :|
Db        111 SSPIPPSHPVSPGTT-------DTHSPSPALPPTQSPESSQRPPL-SSPTGRPDSSTPMR 162

Qy        184 PVCSSKTLQAEVPLSDCVQKASK------PPSSTQIMVKTNMYHNEKVNFHVECKDYVKK 237
              | | :|     | :   :  ||       ||: : : |  :             :
Db        163 PPPSQQTTPPHSPTTPPPEPPSKSSPDSLAPSTLRSLRKRRLSS-------------PQ 208

Qy        238 AKVKINPVQQSRPLL--------SQIHTDAAKENTCYCGAVAKRQEKKGMEPLQ----GH 285
                :||: || |:           ::   | |:  | |: :          |      |
Db        209 GPSTLNPICQSPPVSPPRCDFANRSVYPPWATESPIYVGSSSDGDTPPRQPPTSPISIGS 268

Qy        286 ATPALPF--KETQELLL--------SPLPQEGPGSLAAGESSSLSASTSVSDSSQKKEEH 335
              ::|:      :|  |:|        |   :|   :  |||    :    :|  ||  ::
Db        269 SSPSEGSWGDDTAMLVLLAEIAEEASKNEKECSENNQAGED---NGDNEISKESQVDKDD 325

Qy        336 NYSLFVSDNLGEQPTKCSPEEDEEDEE--------------------------DVDDED 368
              |      ||  ::  : : ||||||:|                            : |||:
Db        326 N------DNKDDEEEQETDEEDEEDDEEDDEEDDEEDDEEDDEEDDEEDDEEDDEEDDEE 379

Qy        369 HDEGFGSEHELSENEEEEEEEEDYEDDKDDDISDTFSEPGYENDSVEDLKE--------- 419
               ||    | :  |:|||:||||| ||| |:|  |       |:|  || ||
Db        380 DDEEDDEEDEEEDEEEDEEEEDEEDDDDEDNEDE------EDDEEEDKKEDEEDGGDGN 433

Qy        420 -VTSISSRKRGKRRYFWEYSEQLTPSQQE 447
               || | :: :        :|  | |||
Db        434 KTLSIQSSQQQQE------PQQQEPQQQE 456


RESULT 11
ADV68154
ID   ADV68154 standard; protein; 1162 AA.
XX
AC   ADV68154;
XX
DT   24-FEB-2005  (first entry)
XX
DE   Kaposi's sarcoma-associated herpesvirus LANA protein.
XX
KW   screening; cytostatic; virucide; gene therapy;
KW   latency-associated nuclear antigen; tumor; viral infection;
KW   herpesvirus type 8 infection.
XX
OS   Human herpesvirus 8.
XX
PN   US2004248081-A1.
XX
PD   09-DEC-2004.
XX
PF   11-JUL-2002; 2002US-00194046.
```

```
XX
PR   01-OCT-1999;   99US-00410399.
XX
PA   (UNMI ) UNIV MICHIGAN.
XX
PI   Robertson ES;
XX
DR   WPI; 2005-038492/04.
DR   N-PSDB; ADV68153.
XX
PT   New composition having a latency-associated nuclear antigen (LANA)
PT   nucleic acid, useful for detecting compounds for treating tumors and
PT   viral infections.
XX
PS   Disclosure; SEQ ID NO 2; 51pp; English.
XX
CC   The invention relates to a composition comprising a latency-associated
CC   nuclear antigen (LANA) nucleic acid having any of 14 fully defined
CC   sequences of 10-61 bp given in the specification (ADV68159-ADV68172). The
CC   methods and compositions of the present invention are useful for
CC   detecting compounds that are agonistic or antagonistic for the binding of
CC   viral genetic material to genomic host DNA, in particular for detecting
CC   compounds for treating tumor and viral infections. This sequence
CC   corresponds to the LANA protein from the Kaposi's sarcoma-associated
CC   herpesvirus (also called human herpesvirus 8). The LANA protein tethers
CC   the viral DNA to the human chromosomal structural protein histone H1.
XX
SQ   Sequence 1162 AA;

  Query Match             8.0%;  Score 190;  DB 9;  Length 1162;
  Best Local Similarity   24.2%;  Pred. No. 3.4e-06;
  Matches   94;  Conservative   50;  Mismatches  137;  Indels  108;  Gaps   14;

Qy        124 SSPYQDEEVISKTPTLAQLNSEDSQSVSDSLYYPDSLFSVKQNPLPSSFPGKKITSRAAA 183
              |||       :|   |       |: | | :|    |  | :: || ||  |:  :|
Db        111 SSPIPPSHPVSPGTT-------DTHSPSPALPPTQSPESSQRPPL-SSPTGRPDSSTPMR 162

Qy        184 PVCSSKTLQAEVPLSDCVQKASK------PPSSTQIMVKTNMYHNEKVNFHVECKDYVKK 237
              |  | :|     | :   :  ||       ||: : : |  :                 :
Db        163 PPPSQQTTPPHSPTTPPPEPPSKSSPDSLAPSTLRSLRKRRLSS--------------PQ 208

Qy        238 AKVKINPVQQSRPLL--------SQIHTDAAKENTCYCGAVAKRQEKKGMEPLQ----GH 285
                :||: || |:           ::   | |:  | |: :       |      |
Db        209 GPSTLNPICQSPPVSPPRCDFANRSVYPPWATESPIYVGSSSDGDTPPRQPPTSPISIGS 268

Qy        286 ATPALPF--KETQELLL--------SPLPQEGPGSLAAGESSSLSASTSVSDSSQKKEEH 335
              ::|:      :|  |:|        |   :|   :  |||    :    :|  ||  ::
Db        269 SSPSEGSWGDDTAMLVLLAEIAEEASKNEKECSENNQAGED---NGDNEISKESQVDKDD 325

Qy        336 NYSLFVSDNLGEQPTKCSPEEDEEDEE--------------------------DVDDED 368
              |      ||  ::  : : ||||||:|                          : |||:
Db        326 N------DNKDDEEEQETDEEDEEDDEEDDEEDDEEDDEEDDEEDDEEDDEEDDEE 379

Qy        369 HDEGFGSEHELSENEEEEEEEEDYEDDKDDDISDTFSEPGYENDSVEDLKE--------- 419
               ||    | :  |:|||:|||||| ||| |:|  |       |:|  || ||
Db        380 DDEEDDEEEDEEEDEEEDEEEEDEDDDDEDNEDE------EDDEEEDKKEDEEDGGDGN 433

Qy        420 -VTSISSRKRGKRRYFWEYSEQLTPSQQE 447
               || | :: :        :|  | |||
Db        434 KTLSIQSSQQQQE------PQQQEPQQQE 456


RESULT 12
ADE61623
ID   ADE61623 standard; protein; 712 AA.
XX
AC   ADE61623;
XX
DT   29-JAN-2004  (first entry)
XX
DE   Rat Protein P13383, SEQ ID NO 7544.
XX
KW   Rat; pain; neuronal tissue; gene therapy; spinal segmental nerve injury;
KW   chronic constriction injury; CCI; spared nerve injury; SNI; Chung.
XX
OS   Rattus norvegicus.
XX
PN   WO2003016475-A2.
XX
PD   27-FEB-2003.
XX
PF   14-AUG-2002; 2002WO-US025765.
XX
PR   14-AUG-2001; 2001US-0312147P.
PR   01-NOV-2001; 2001US-0346382P.
PR   26-NOV-2001; 2001US-0333347P.
XX
PA   (GEHO ) GEN HOSPITAL CORP.
PA   (FARB ) BAYER AG.
XX
PI   Woolf C,  D'urso D,  Befort K,  Costigan M;
XX
DR   WPI; 2003-268312/26.
DR   GENBANK; P13383.
XX
PT   New composition comprising two or more isolated polypeptides, useful for
PT   preparing a medicament for treating pain in an animal.
XX
PS   Claim 1; Page; 1017pp; English.
XX
CC   The invention discloses a composition comprising two or more isolated rat
CC   or human polynucleotides or a polynucleotide which represents a fragment,
CC   derivative or allelic variation of the nucleic acid sequence. Also
CC   claimed are a vector comprising the novel polynucleotide, a host cell
CC   comprising the vector, a method for identifying a nucleotide sequence
CC   which is differentially regulated in an animal subjected to pain and a
CC   kit to perform the method, an array, a method for identifying an agent
CC   that increases or decreases the expression of the polynucleotide sequence
CC   that is differentially expressed in neuronal tissue of a first animal
CC   subjected to pain, a method for identifying a compound which regulates
CC   the expression of a polynucleotide sequence which is differentially
CC   expressed in an animal subjected to pain, a method for identifying a
CC   compound that regulates the activity of one or more of the
CC   polynucleotides, a method for producing a pharmaceutical composition, a
CC   method for identifying a compound or small molecule that regulates the
CC   activity in an animal of one or more of the polypeptides given in the
CC   specification, a method for identifying a compound useful in treating
CC   pain and a pharmaceutical composition comprising the one or more
CC   polypeptides or their antibodies. The polynucleotide or the compound that
CC   modulates its activity is useful for preparing a medicament for treating
CC   pain (e.g. spinal segmental nerve injury (Chung), chronic constriction
CC   injury (CCI) and spared nerve injury (SNI)) in an animal (e.g. gene
CC   therapy). The sequence presented is a rat protein (shown in Table 2 of
CC   the specification) which is differentially expressed during pain. Note:
CC   The sequence data for this patent did not form part of the printed
CC   specification, but was obtained in electronic form directly from WIPO at
CC   ftp.wipo.int/pub/published_pct_sequences.
XX
SQ   Sequence 712 AA;

  Query Match             7.8%;  Score 185.5;  DB 7;  Length 712;
  Best Local Similarity   23.3%;  Pred. No. 3.9e-06;
  Matches   84;  Conservative   34;  Mismatches  134;  Indels  109;  Gaps   11;

Qy        129 DEEVISKTPTLAQLNSEDSQSVSDSLYYPDSLFSVKQNPLPSSFPGKKITSRAAAPVCSS 188
              : : ::  |   : :||| :   |     |      ::  :|    ||| |:  |  |  |
Db         12 ESKKMAPPPKEVEEDSEDEEMSEDE----DDSSGEEEVVIPQK-KGKKATTTPAKKVVVS 66

Qy        189 KTLQAEVPLSDCVQKASKPPSSTQIMVKTNMYHNEKVNFHVECKDYVKKAKVKINPVQQS 248
              :| :| ||     :||:  |                |        |  |  |||   |
Db         67 QTKKAAVPTP--AKKAAVTPG--------------KKAAATPAKKAVTPAKVVPTP---- 106

Qy        249 RPLLSQIHTDAAKENTCYCGAVAKRQEKKGMEPLQGHATPALPFKETQELLLSPLPQEGP 308
                          |:      |:    |||         |||   |
Db        107 -----------GKKGAAQAKALVPTPGKKGA------VTPAKGAKN-------------- 135
```

```
Qy        309 GSLAAGESSSLSASTSVSDSSQKKEEHNYSLFVSDNLGEQPTKCSP------EEDEEDEE 362
              |  |  | |          | | : |:           | :| | :|      :|||||::
Db        136 GKNAKKEDSDEDEDEEDEDDSDEDEDEEDEFEPPVVKGVKPAKAAPAAPASEDEDEEDDD 195

Qy        363 DVDDEDHDE--------------------------------GFGSEHELSENEEEEEEE 389
              | ||:| ||                                | |  |::|:|||:
Db        196 DEDDDDDDEEEEEEDDSEEEVMEITPAKGKKTPAKVVPVKAKSVAEEEEDDEDDEDEEED 255

Qy        390 EDYEDDKDDDISDTFSE-------PGYENDSVEDLKEVTSISSRKRGKRRYFWEYSEQLT 442
              || ||::||:  |   |       ||     :   ||      :|        | ||  |
Db        256 EDEEDEEDDEDEDEEEEEEPVKAAPGKRKKEMTKQKEAPEAKKQK-------IEGSEPTT 308

Qy        443 P 443
              |
Db        309 P 309
```

RESULT 13
ADD46090
ID ADD46090 standard; protein; 712 AA.
XX
AC ADD46090;
XX
DT 29-JAN-2004 (first entry)
XX
DE Rat Protein P13383, SEQ ID NO 11765.
XX
KW Rat; pain; neuronal tissue; gene therapy; spinal segmental nerve injury;
KW chronic constriction injury; CCI; spared nerve injury; SNI; Chung.
XX
OS Rattus norvegicus.
XX
PN WO2003016475-A2.
XX
PD 27-FEB-2003.
XX
PF 14-AUG-2002; 2002WO-US025765.
XX
PR 14-AUG-2001; 2001US-0312147P.
PR 01-NOV-2001; 2001US-0346382P.
PR 26-NOV-2001; 2001US-0333347P.
XX
PA (GEHO) GEN HOSPITAL CORP.
PA (FARB) BAYER AG.
XX
PI Woolf C, D'urso D, Befort K, Costigan M;
XX
DR WPI; 2003-268312/26.
DR GENBANK; P13383.
XX
PT New composition comprising two or more isolated polypeptides, useful for
PT preparing a medicament for treating pain in an animal.
XX
PS Claim 1; Page; 1017pp; English.
XX
CC The invention discloses a composition comprising two or more isolated rat
CC or human polynucleotides or a polynucleotide which represents a fragment,
CC derivative or allelic variation of the nucleic acid sequence. Also
CC claimed are a vector comprising the novel polynucleotide, a host cell
CC comprising the vector, a method for identifying a nucleotide sequence
CC which is differentially regulated in an animal subjected to pain and a
CC kit to perform the method, an array, a method for identifying an agent
CC that increases or decreases the expression of the polynucleotide sequence
CC that is differentially expressed in neuronal tissue of a first animal
CC subjected to pain, a method for identifying a compound which regulates
CC the expression of a polynucleotide sequence which is differentially
CC expressed in an animal subjected to pain, a method for identifying a
CC compound that regulates the activity of one or more of the
CC polynucleotides, a method for producing a pharmaceutical composition, a
CC method for identifying a compound or small molecule that regulates the
CC activity in an animal of one or more of the polypeptides given in the
CC specification, a method for identifying a compound useful in treating
CC pain and a pharmaceutical composition comprising the one or more
CC polypeptides or their antibodies. The polynucleotide or the compound that
CC modulates its activity is useful for preparing a medicament for treating
CC pain (e.g. spinal segmental nerve injury (Chung), chronic constriction
CC injury (CCI) and spared nerve injury (SNI)) in an animal (e.g. gene
CC therapy). The sequence presented is a rat protein (shown in Table 2 of
CC the specification) which is differentially expressed during pain. Note:
CC The sequence data for this patent did not form part of the printed
CC specification, but was obtained in electronic form directly from WIPO at
CC ftp.wipo.int/pub/published_pct_sequences.
XX
SQ Sequence 712 AA;

Query Match 7.8%; Score 185.5; DB 7; Length 712;
Best Local Similarity 23.3%; Pred. No. 3.9e-06;
Matches 84; Conservative 34; Mismatches 134; Indels 109; Gaps 11;

```
Qy        129 DEEVISKTPTLAQLNSEDSQSVSDSLYYPDSLFSVKQNPLPSSFPGKKITSRAAAPVCSS 188
              : : ::  |   : :||| :   |     |     ::  :|    ||| |:  |  |  |
Db         12 ESKKMAPPPKEVEEDSEDEEMSEDE----DDSSGEEEVVIPQK-KGKKATTTPAKKVVVS 66

Qy        189 KTLQAEVPLSDCVQKASKPPSSTQIMVKTNMYHNEKVNFHVECKDYVKKAKVKINPVQQS 248
              :| :| ||    :||:  |                |       |  |  |||   |
Db         67 QTKKAAVPTP--AKKAAVTPG-------------KKAAATPAKKAVTPAKVVPTP---- 106

Qy        249 RPLLSQIHTDAAKENTCYCGAVAKRQEKKGMEPLQGHATPALPFKETQELLLSPLPQEGP 308
                          |:      |:     |||       |||   |
Db        107 -----------GKKGAAQAKALVPTPGKKGA------VTPAKGAKN-------------- 135

Qy        309 GSLAAGESSSLSASTSVSDSSQKKEEHNYSLFVSDNLGEQPTKCSP------EEDEEDEE 362
              |  |  | |          | | : |:           | :| | :|      :|||||::
Db        136 GKNAKKEDSDEDEDEEDEDDSDEDEDEEDEFEPPVVKGVKPAKAAPAAPASEDEDEEDDD 195

Qy        363 DVDDEDHDE--------------------------------GFGSEHELSENEEEEEEE 389
              | ||:| ||                                | |  |::|:|||:
Db        196 DEDDDDDDEEEEEEDDSEEEVMEITPAKGKKTPAKVVPVKAKSVAEEEEDDEDDEDEEED 255

Qy        390 EDYEDDKDDDISDTFSE-------PGYENDSVEDLKEVTSISSRKRGKRRYFWEYSEQLT 442
              || ||::||:  |   |       ||     :   ||      :|        | ||  |
Db        256 EDEEDEEDDEDEDEEEEEEPVKAAPGKRKKEMTKQKEAPEAKKQK-------IEGSEPTT 308

Qy        443 P 443
              |
Db        309 P 309
```

RESULT 14
ADE57828
ID ADE57828 standard; protein; 712 AA.
XX
AC ADE57828;
XX
DT 29-JAN-2004 (first entry)
XX
DE Rat Protein AAA41732, SEQ ID NO 3693.
XX
KW Rat; pain; neuronal tissue; gene therapy; spinal segmental nerve injury;
KW chronic constriction injury; CCI; spared nerve injury; SNI; Chung.
XX
OS Rattus norvegicus.
XX
PN WO2003016475-A2.
XX
PD 27-FEB-2003.
XX
PF 14-AUG-2002; 2002WO-US025765.
XX
PR 14-AUG-2001; 2001US-0312147P.
PR 01-NOV-2001; 2001US-0346382P.
PR 26-NOV-2001; 2001US-0333347P.
XX
PA (GEHO) GEN HOSPITAL CORP.
PA (FARB) BAYER AG.
XX

```
PI   Woolf C,  D'urso D,  Befort K,  Costigan M;
XX
DR   WPI; 2003-268312/26.
DR   GENBANK; AAA41732.
XX
PT   New composition comprising two or more isolated polypeptides, useful for
PT   preparing a medicament for treating pain in an animal.
XX
PS   Claim 1; Page; 1017pp; English.
XX
CC   The invention discloses a composition comprising two or more isolated rat
CC   or human polynucleotides or a polynucleotide which represents a fragment,
CC   derivative or allelic variation of the nucleic acid sequence. Also
CC   claimed are a vector comprising the novel polynucleotide, a host cell
CC   comprising the vector, a method for identifying a nucleotide sequence
CC   which is differentially regulated in an animal subjected to pain and a
CC   kit to perform the method, an array, a method for identifying an agent
CC   that increases or decreases the expression of the polynucleotide sequence
CC   that is differentially expressed in neuronal tissue of a first animal
CC   subjected to pain, a method for identifying a compound which regulates
CC   the expression of a polynucleotide sequence which is differentially
CC   expressed in an animal subjected to pain, a method for identifying a
CC   compound that regulates the activity of one or more of the
CC   polynucleotides, a method for producing a pharmaceutical composition, a
CC   method for identifying a compound or small molecule that regulates the
CC   activity in an animal of one or more of the polypeptides given in the
CC   specification, a method for identifying a compound useful in treating
CC   pain and a pharmaceutical composition comprising the one or more
CC   polypeptides or their antibodies. The polynucleotide or the compound that
CC   modulates its activity is useful for preparing a medicament for treating
CC   pain (e.g. spinal segmental nerve injury (Chung), chronic constriction
CC   injury (CCI) and spared nerve injury (SNI)) in an animal (e.g. gene
CC   therapy). The sequence presented is a rat protein (shown in Table 2 of
CC   the specification) which is differentially expressed during pain. Note:
CC   The sequence data for this patent did not form part of the printed
CC   specification, but was obtained in electronic form directly from WIPO at
CC   ftp.wipo.int/pub/published_pct_sequences.
XX
SQ   Sequence 712 AA;

  Query Match             7.8%;  Score 185.5;  DB 7;  Length 712;
  Best Local Similarity   23.3%;  Pred. No. 3.9e-06;
  Matches   84;  Conservative   34;  Mismatches  134;  Indels  109;  Gaps   11;

Qy        129 DEEVISKTPTLAQLNSEDSQSVSDSLYYPDSLFSVKQNPLPSSFPGKKITSRAAAPVCSS 188
              : : ::  |   : :||| :   |     |     ::  :|    ||| |:  |  |  |
Db         12 ESKKMAPPPKEVEEDSEDEEMSEDE----DDSSGEEEVVIPQK-KGKKATTTPAKKVVVS 66

Qy        189 KTLQAEVPLSDCVQKASKPPSSTQIMVKTNMYHNEKVNFHVECKDYVKKAKVKINPVQQS 248
              :| :| ||    :||: |                |   |  |  |||   |
Db         67 QTKKAAVPTP--AKKAAVTPG-------------KKAAATPAKKAVTPAKVVPTP---- 106

Qy        249 RPLLSQIHTDAAKENTCYCGAVAKRQEKKGMEPLQGHATPALPFKETQELLLSPLPQEGP 308
                          |:     |:     |||        |||   |
Db        107 -----------GKKGAAQAKALVPTPGKKGA------VTPAKGAKN------------- 135

Qy        309 GSLAAGESSSLSASTSVSDSSQKKEEHNYSLFVSDNLGEQPTKCSP------EEDEEDEE 362
              |  |  | |          | | : |:          | :| | :|      :|||||::
Db        136 GKNAKKEDSDEDEDEEDEDDSDEDEDEEDEFEPPVVKGVKPAKAAPAAPASEDEDEEDDD 195

Qy        363 DVDDEDHDE--------------------------------GFGSEHELSENEEEEEEE 389
              | ||:| ||                                |  |::|:|||:
Db        196 DEDDDDDDEEEEEEDDSEEEVMEITPAKGKKTPAKVVPVKAKSVAEEEEDDEDDEDEEED 255

Qy        390 EDYEDDKDDDISDTFSE-------PGYENDSVEDLKEVTSISSRKRGKRRYFWEYSEQLT 442
              || ||::||:  |   |       ||     :  ||      :|        | ||  |
Db        256 EDEEDEEDDEDEDEEEEEPVKAAPGKRKKEMTKQKEAPEAKKQK-------IEGSEPTT 308

Qy        443 P 443
              |
Db        309 P 309
```

```
RESULT 15
ADE57830
ID   ADE57830 standard; protein; 712 AA.
XX
AC   ADE57830;
XX
DT   29-JAN-2004  (first entry)
XX
DE   Rat Protein AAA41732, SEQ ID NO 3695.
XX
KW   Rat; pain; neuronal tissue; gene therapy; spinal segmental nerve injury;
KW   chronic constriction injury; CCI; spared nerve injury; SNI; Chung.
XX
OS   Rattus norvegicus.
XX
PN   WO2003016475-A2.
XX
PD   27-FEB-2003.
XX
PF   14-AUG-2002; 2002WO-US025765.
XX
PR   14-AUG-2001; 2001US-0312147P.
PR   01-NOV-2001; 2001US-0346382P.
PR   26-NOV-2001; 2001US-0333347P.
XX
PA   (GEHO ) GEN HOSPITAL CORP.
PA   (FARB ) BAYER AG.
XX
PI   Woolf C,  D'urso D,  Befort K,  Costigan M;
XX
DR   WPI; 2003-268312/26.
DR   GENBANK; AAA41732.
XX
PT   New composition comprising two or more isolated polypeptides, useful for
PT   preparing a medicament for treating pain in an animal.
XX
PS   Claim 1; Page; 1017pp; English.
XX
CC   The invention discloses a composition comprising two or more isolated rat
CC   or human polynucleotides or a polynucleotide which represents a fragment,
CC   derivative or allelic variation of the nucleic acid sequence. Also
CC   claimed are a vector comprising the novel polynucleotide, a host cell
CC   comprising the vector, a method for identifying a nucleotide sequence
CC   which is differentially regulated in an animal subjected to pain and a
CC   kit to perform the method, an array, a method for identifying an agent
CC   that increases or decreases the expression of the polynucleotide sequence
CC   that is differentially expressed in neuronal tissue of a first animal
CC   subjected to pain, a method for identifying a compound which regulates
CC   the expression of a polynucleotide sequence which is differentially
CC   expressed in an animal subjected to pain, a method for identifying a
CC   compound that regulates the activity of one or more of the
CC   polynucleotides, a method for producing a pharmaceutical composition, a
CC   method for identifying a compound or small molecule that regulates the
CC   activity in an animal of one or more of the polypeptides given in the
CC   specification, a method for identifying a compound useful in treating
CC   pain and a pharmaceutical composition comprising the one or more
CC   polypeptides or their antibodies. The polynucleotide or the compound that
CC   modulates its activity is useful for preparing a medicament for treating
CC   pain (e.g. spinal segmental nerve injury (Chung), chronic constriction
CC   injury (CCI) and spared nerve injury (SNI)) in an animal (e.g. gene
CC   therapy). The sequence presented is a rat protein (shown in Table 2 of
CC   the specification) which is differentially expressed during pain. Note:
CC   The sequence data for this patent did not form part of the printed
CC   specification, but was obtained in electronic form directly from WIPO at
CC   ftp.wipo.int/pub/published_pct_sequences.
XX
SQ   Sequence 712 AA;

  Query Match             7.8%;  Score 185.5;  DB 7;  Length 712;
  Best Local Similarity   23.3%;  Pred. No. 3.9e-06;
  Matches   84;  Conservative   34;  Mismatches  134;  Indels  109;  Gaps   11;

Qy        129 DEEVISKTPTLAQLNSEDSQSVSDSLYYPDSLFSVKQNPLPSSFPGKKITSRAAAPVCSS 188
```

```
                : : ::  |   : :||| :   |     |     ::  :|    ||| |:  |  |  |
Db         12 ESKKMAPPPKEVEEDSEDEEMSEDE----DDSSGEEEVVIPQK-KGKKATTTPAKKVVVS 66

Qy        189 KTLQAEVPLSDCVQKASKPPSSTQIMVKTNMYHNEKVNFHVECKDYVKKAKVKINPVQQS 248
              :| :| ||    :||:  |               |       |  |  |||   |
Db         67 QTKKAAVPTP--AKKAAVTPG-------------KKAAATPAKKAVTPAKVVPTP---- 106

Qy        249 RPLLSQIHTDAAKENTCYCGAVAKRQEKKGMEPLQGHATPALPFKETQELLLSPLPQEGP 308
                        |:      |:    |||        |||   |
Db        107 -----------GKKGAAQAKALVPTPGKKGA------VTPAKGAKN-------------- 135

Qy        309 GSLAAGESSSLSASTSVSDSSQKKEEHNYSLFVSDNLGEQPTKCSP------EEDEEDEE 362
              |  |  | |          | | : |:           | :|'| :|      :|||||::
Db        136 GKNAKKEDSDEDEDEEDEDDSDEDEDEEDEFEPPVVKGVKPAKAAPAAPASEDEDEEDDD 195

Qy        363 DVDDEDHDE--------------------------------GFGSEHELSENEEEEEEE 389
              | ||:| ||                                | |  |::|:|||:
Db        196 DEDDDDDDEEEEEEDDSEEEVMEITPAKGKKTPAKVVPVKAKSVAEEEEDDEDDEDEEED 255

Qy        390 EDYEDDKDDDISDTFSE-------PGYENDSVEDLKEVTSISSRKRGKRRYFWEYSEQLT 442
              || ||::||:  |   |       ||     :   ||      :|        | ||  |
Db        256 EDEEDEEDDEDEDEEEEEEPVKAAPGKRKKEMTKQKEAPEAKKQK-------IEGSEPTT 308

Qy        443 P 443
              |
Db        309 P 309


Search completed: February 28, 2006, 08:45:06
Job time : 155.45 secs
```

This Page Blank (uspto)

GenCore version 5.1.7

OM protein - protein search, using sw model

Run on: February 28, 2006, 08:53:26 ; Search time 38.5497 Seconds
(without alignments)
960.804 Million cell updates/sec

Title: US-10-717-665A-44_COPY_1_448
Perfect score: 2364
Sequence: 1 MPQPSVSGMDPPFGDAFRSH.....GKRRYFWEYSEQLTPSQQER 448

Scoring table: BLOSUM62
Gapop 10.0 , Gapext 0.5

Searched: 572060 seqs, 82675679 residues

Total number of hits satisfying chosen parameters: 572060

Minimum DB seq length: 0
Maximum DB seq length: 2000000000

Post-processing: Minimum Match 0%
Maximum Match 100%
Listing first 45 summaries

Database : Issued_Patents_AA:*
1: /cgn2_6/ptodata/1/iaa/5_COMB.pep:*
2: /cgn2_6/ptodata/1/iaa/6_COMB.pep:*
3: /cgn2_6/ptodata/1/iaa/H_COMB.pep:*
4: /cgn2_6/ptodata/1/iaa/PCTUS_COMB.pep:*
5: /cgn2_6/ptodata/1/iaa/RE_COMB.pep:*
6: /cgn2_6/ptodata/1/iaa/backfiles1.pep:*

Pred. No. is the number of results predicted by chance to have a score greater than or equal to the score of the result being printed, and is derived by analysis of the total score distribution.

SUMMARIES

Result No.	Score	% Query Match	Length	DB	ID	Description
1	2364	100.0	639	2	US-10-164-595-44	Sequence 44, Appl
2	190	8.0	1162	1	US-08-728-323A-2	Sequence 2, Appli
3	190	8.0	1162	2	US-09-298-568-2	Sequence 2, Appli
4	190	8.0	1162	2	US-09-410-399-2	Sequence 2, Appli
5	190	8.0	1162	2	US-09-894-273-2	Sequence 2, Appli
6	165	7.0	706	2	US-09-538-092-957	Sequence 957, App
7	165	7.0	747	2	US-09-949-016-10040	Sequence 10040, A
8	161.5	6.8	709	2	US-09-248-796A-19045	Sequence 19045, A
9	161	6.8	831	1	US-09-047-026A-4	Sequence 4, Appli
10	160	6.8	1972	2	US-09-418-710-21	Sequence 21, Appl
11	160	6.8	1972	2	US-09-839-479-21	Sequence 21, Appl
12	157.5	6.7	738	2	US-09-248-796A-16666	Sequence 16666, A
13	157	6.6	1969	2	US-09-418-710-72	Sequence 72, Appl
14	157	6.6	1969	2	US-09-839-479-71	Sequence 71, Appl
15	155.5	6.6	687	2	US-10-104-047-2651	Sequence 2651, Ap
16	151	6.4	311	2	US-09-902-540-10544	Sequence 10544, A
17	151	6.4	905	1	US-08-574-959A-9	Sequence 9, Appli
18	151	6.4	905	2	US-09-357-014-9	Sequence 9, Appli
19	151	6.4	1135	1	US-08-574-959A-7	Sequence 7, Appli
20	151	6.4	1135	2	US-09-357-014-7	Sequence 7, Appli
21	150.5	6.4	714	1	US-08-990-114-3	Sequence 3, Appli
22	150.5	6.4	714	2	US-09-241-333-3	Sequence 3, Appli
23	150.5	6.4	740	2	US-09-022-983-5	Sequence 5, Appli
24	150.5	6.4	742	2	US-09-949-016-11569	Sequence 11569, A
25	149.5	6.3	542	2	US-08-935-855-22	Sequence 22, Appl
26	148.5	6.3	764	1	US-08-375-300-4	Sequence 4, Appli
27	148.5	6.3	764	2	US-09-177-431-4	Sequence 4, Appli
28	148.5	6.3	764	4	PCT-US95-16930-4	Sequence 4, Appli
29	148.5	6.3	1089	1	US-08-375-300-2	Sequence 2, Appli
30	148.5	6.3	1089	2	US-09-177-431-2	Sequence 2, Appli
31	148.5	6.3	1089	4	PCT-US95-16930-2	Sequence 2, Appli
32	146.5	6.2	1089	2	US-09-949-016-10326	Sequence 10326, A
33	146	6.2	933	2	US-08-293-728-2	Sequence 2, Appli
34	146	6.2	933	2	US-09-421-868-2	Sequence 2, Appli
35	146	6.2	936	2	US-08-956-171E-5249	Sequence 5249, Ap
36	146	6.2	936	2	US-08-781-986A-5249	Sequence 5249, Ap
37	145	6.1	1085	1	US-08-431-080-28	Sequence 28, Appl
38	145	6.1	1085	1	US-08-938-534-28	Sequence 28, Appl
39	145	6.1	1085	2	US-09-345-294-28	Sequence 28, Appl
40	143	6.0	392	1	US-08-822-701-2	Sequence 2, Appli
41	143	6.0	392	2	US-08-935-855-2	Sequence 2, Appli
42	142.5	6.0	279	2	US-09-699-266A-7	Sequence 7, Appli
43	142	6.0	226	1	US-08-431-080-26	Sequence 26, Appl
44	142	6.0	226	1	US-08-938-534-26	Sequence 26, Appl
45	142	6.0	226	2	US-09-345-294-26	Sequence 26, Appl

ALIGNMENTS

RESULT 1
US-10-164-595-44
; Sequence 44, Application US/10164595
; Patent No. 6657054
; GENERAL INFORMATION:
; APPLICANT: OriGene Technologies, Inc
; TITLE OF INVENTION: Regulated Angiogenesis Genes and Polypeptides
; FILE REFERENCE: 1U 103 R1
; CURRENT APPLICATION NUMBER: US/10/164,595
; CURRENT FILING DATE: 2002-06-10
; NUMBER OF SEQ ID NOS: 80
; SOFTWARE: PatentIn version 3.1
; SEQ ID NO 44
; LENGTH: 639
; TYPE: PRT
; ORGANISM: Homo sapiens
US-10-164-595-44

Query Match 100.0%; Score 2364; DB 2; Length 639;
Best Local Similarity 100.0%; Pred. No. 4.1e-200;
Matches 448; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

```
Qy          1 MPQPSVSGMDPPFGDAFRSHTFSEQTLMSTDLLANSSDPDFMYELDREMNYQQNPRDNFL 60
              ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db          1 MPQPSVSGMDPPFGDAFRSHTFSEQTLMSTDLLANSSDPDFMYELDREMNYQQNPRDNFL 60

Qy         61 SLEDCKDIENLESFTDVLDNEGALTSNWEQWDTYCEDLTKYTKLTSCDIWGTKEVDYLGL 120
              ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db         61 SLEDCKDIENLESFTDVLDNEGALTSNWEQWDTYCEDLTKYTKLTSCDIWGTKEVDYLGL 120

Qy        121 DDFSSPYQDEEVISKTPTLAQLNSEDSQSVSDSLYYPDSLFSVKQNPLPSSFPGKKITSR 180
              ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db        121 DDFSSPYQDEEVISKTPTLAQLNSEDSQSVSDSLYYPDSLFSVKQNPLPSSFPGKKITSR 180

Qy        181 AAAPVCSSKTLQAEVPLSDCVQKASKPPSSTQIMVKTNMYHNEKVNFHVECKDYVKKAKV 240
              ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db        181 AAAPVCSSKTLQAEVPLSDCVQKASKPPSSTQIMVKTNMYHNEKVNFHVECKDYVKKAKV 240

Qy        241 KINPVQQSRPLLSQIHTDAAKENTCYCGAVAKRQEKKGMEPLQGHATPALPFKETQELLL 300
              ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db        241 KINPVQQSRPLLSQIHTDAAKENTCYCGAVAKRQEKKGMEPLQGHATPALPFKETQELLL 300

Qy        301 SPLPQEGPGSLAAGESSSLSASTSVSDSSQKKEEHNYSLFVSDNLGEQPTKCSPEEDEED 360
              ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db        301 SPLPQEGPGSLAAGESSSLSASTSVSDSSQKKEEHNYSLFVSDNLGEQPTKCSPEEDEED 360

Qy        361 EEDVDDEDHDEGFGSEHELSENEEEEEEEDYEDDKDDDISDTFSEPGYENDSVEDLKEV 420
              ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db        361 EEDVDDEDHDEGFGSEHELSENEEEEEEEDYEDDKDDDISDTFSEPGYENDSVEDLKEV 420
```

```
Qy        421 TSISSRKRGKRRYFWEYSEQLTPSQQER 448
              ||||||||||||||||||||||||||||||
Db        421 TSISSRKRGKRRYFWEYSEQLTPSQQER 448
```

RESULT 2
US-08-728-323A-2
; Sequence 2, Application US/08728323A
; Patent No. 5948676
; GENERAL INFORMATION:
; APPLICANT: Chang, Yuan
; APPLICANT: Bohenzky, Roy A.
; APPLICANT: Russo, James J.
; APPLICANT: Edelman, Isidore S.
; APPLICANT: Moore, Patrick S.
; TITLE OF INVENTION: Immediate Early Protein From Kaposi's
; TITLE OF INVENTION: Sarcoma-Associated Herpesvirus, DNA
; TITLE OF INVENTION: Encoding Same And Uses Thereof
; NUMBER OF SEQUENCES: 21
; CORRESPONDENCE ADDRESS:
; ADDRESSEE: Cooper & Dunham LLP
; STREET: 1185 Avenue of the Americas
; CITY: New York
; STATE: New York
; COUNTRY: U.S.A.
; ZIP: 10036
; COMPUTER READABLE FORM:
; MEDIUM TYPE: Floppy disk
; COMPUTER: IBM PC compatible
; OPERATING SYSTEM: PC-DOS/MS-DOS
; SOFTWARE: PatentIn Release #1.0, Version #1.30
; CURRENT APPLICATION DATA:
; APPLICATION NUMBER: US/08/728,323A
; FILING DATE:
; CLASSIFICATION: 435
; ATTORNEY/AGENT INFORMATION:
; NAME: White, John P.
; REGISTRATION NUMBER: 28,678
; REFERENCE/DOCKET NUMBER: 0575/52268/JPW/MSC/SKS
; TELECOMMUNICATION INFORMATION:
; TELEPHONE: 212-278-0400
; TELEFAX: 212-391-0525
; INFORMATION FOR SEQ ID NO: 2:
; SEQUENCE CHARACTERISTICS:
; LENGTH: 1162 amino acids
; TYPE: amino acid
; TOPOLOGY: linear
; MOLECULE TYPE: protein
US-08-728-323A-2

Query Match 8.0%; Score 190; DB 1; Length 1162;
Best Local Similarity 24.2%; Pred. No. 1.4e-07;
Matches 94; Conservative 50; Mismatches 137; Indels 108; Gaps 14;

```
Qy        124 SSPYQDEEVISKTPTLAQLNSEDSQSVSDSLYYPDSLFSVKQNPLPSSFPGKKITSRAAA 183
              |||      :|   |       |: | | :|    |  | :: || ||  |:  :|
Db        111 SSPIPPSHPVSPGTT-------DTHSPSPALPPTQSPESSQRPPL-SSPTGRPDSSTPMR 162

Qy        184 PVCSSKTLQAEVPLSDCVQKASK------PPSSTQIMVKTNMYHNEKVNFHVECKDYVKK 237
              | | :|     | :   :  ||       ||: : : |  :                 :
Db        163 PPPSQQTTPPHSPTTPPPEPPSKSSPDSLAPSTLRSLRKRRLSS-------------PQ 208

Qy        238 AKVKINPVQQSRPLL--------SQIHTDAAKENTCYCGAVAKRQEKKGMEPLQ----GH 285
                :||: || |:          ::   | |:  | |: :          |      |
Db        209 GPSTLNPICQSPPVSPPRCDFANRSVYPPWATESPIYVGSSSDGDTPPRQPPTSPISIGS 268

Qy        286 ATPALPF--KETQELLL--------SPLPQEGPGSLAAGESSSLSASTSVSDSSQKKEEH 335
              ::|:      :|  |:|        |   :|   :  |||    :    :|  ||  ::
Db        269 SSPSEGSWGDDTAMLVLLAEIAEEASKNEKECSENNQAGED---NGDNEISKESQVDKDD 325

Qy        336 NYSLFVSDNLGEQPTKCSPEEDEEDEE-------------------------DVDDED 368
              |      ||  ::  : : ||||||:|                         : |||:
Db        326 N------DNKDDEEEQETDEEDEEDDEEDDEEDDEEDDEEDDEEDDEEDDEEDDEE 379

Qy        369 HDEGFGSEHELSENEEEEEEEEDYEDDKDDDISDTFSEPGYENDSVEDLKE--------- 419
              ||    | :  |:|||:||||| ||| |:|  |       |:|  || ||
Db        380 DDEEDDEEEDEEEDEEEDEEEEDEEDDDDEDNEDE------EDDEEEDKKEDEEDGGDGN 433

Qy        420 -VTSISSRKRGKRRYFWEYSEQLTPSQQE 447
               || | :: :        :|  | |||
Db        434 KTLSIQSSQQQQE------PQQQEPQQQE 456

RESULT 3
US-09-298-568-2
; Sequence 2, Application US/09298568
; Patent No. 6322792
; GENERAL INFORMATION:
; APPLICANT: Kieff, Elliott D.
; APPLICANT: Ballestas, Mary E.
; APPLICANT: Kaye, Kenneth M.
; TITLE OF INVENTION: RHADINO VIRUS LANA ACTS IN TRANS ON A UNIT OF RHADINO
; TITLE OF INVENTION: VIRUS DNA TO MEDIATE EFFICIENT EPISOME PERSISTENCE
; FILE REFERENCE: 16412-10001R
; CURRENT APPLICATION NUMBER: US/09/298,568
; CURRENT FILING DATE: 1999-04-21
; EARLIER APPLICATION NUMBER: US 60/109,422
; EARLIER FILING DATE: 1998-11-19
; NUMBER OF SEQ ID NOS: 3
; SOFTWARE: PatentIn Ver. 2.0
; SEQ ID NO 2
; LENGTH: 1162
; TYPE: PRT
; ORGANISM: Kaposi's sarcoma-associated herpesvirus
US-09-298-568-2

Query Match 8.0%; Score 190; DB 2; Length 1162;
Best Local Similarity 24.2%; Pred. No. 1.4e-07;
Matches 94; Conservative 50; Mismatches 137; Indels 108; Gaps 14;

```
Qy   124 SSPYQDEEVISKTPTLAQLNSEDSQSVSDSLYYPDSLFSVKQNPLPSSFPGKKITSRAAA 183
              ||| :| | |: | | :| | | :: || || |: :|
Db   111 SSPIPPSHPVSPGTT-------DTHSPSPALPPTQSPESSQRPPL-SSPTGRPDSSTPMR 162

Qy   184 PVCSSKTLQAEVPLSDCVQKASK------PPSSTQIMVKTNMYHNEKVNFHVECKDYVKK 237
              | | :| | :   : || ||: : : | :   :
Db   163 PPPSQQTTPPHSPTTPPPEPPSKSSPDSLAPSTLRSLRKRRLSS-------------PQ 208

Qy   238 AKVKINPVQQSRPLL--------SQIHTDAAKENTCYCGAVAKRQEKKGMEPLQ----GH 285
                :||: || |:   :: | |: | |: : | |
Db   209 GPSTLNPICQSPPVSPPRCDFANRSVYPPWATESPIYVGSSSDGDTPPRQPPTSPISIGS 268

Qy   286 ATPALPF--KETQELLL--------SPLPQEGPGSLAAGESSSLSASTSVSDSSQKKEEH 335
              ::|:   :| |:| | :| : ||| :   :| || ::
Db   269 SSPSEGSWGDDTAMLVLLAEIAEEASKNEKECSENNQAGED---NGDNEISKESQVDKDD 325

Qy   336 NYSLFVSDNLGEQPTKCSPEEDEEDEE-------------------------DVDDED 368
              | || :: : : ||||||:| : |||:
Db   326 N------DNKDDEEEQETDEEDEEDDEEDDEEDDEEDDEEDDEEDDEEDDEEDDEE 379

Qy   369 HDEGFGSEHELSENEEEEEEEEDYEDDKDDDISDTFSEPGYENDSVEDLKE--------- 419
              || | : |:|||:||||| ||| |:| | |:| || ||
Db   380 DDEEDDEEEDEEEDEEEDEEEDEEDDEEDDEEDDEEDDEEDDEEDDEEDDEEDDEE 433

Qy   420 -VTSISSRKRGKRRYFWEYSEQLTPSQQE 447
               || | :: :   :| | |||
Db   434 KTLSIQSSQQQQE------PQQQEPQQQE 456

RESULT 4
US-09-410-399-2
; Sequence 2, Application US/09410399
; Patent No. 6482587
; GENERAL INFORMATION:
; APPLICANT: Robertson, Erle S.
; APPLICANT: Cotter, Murray A.

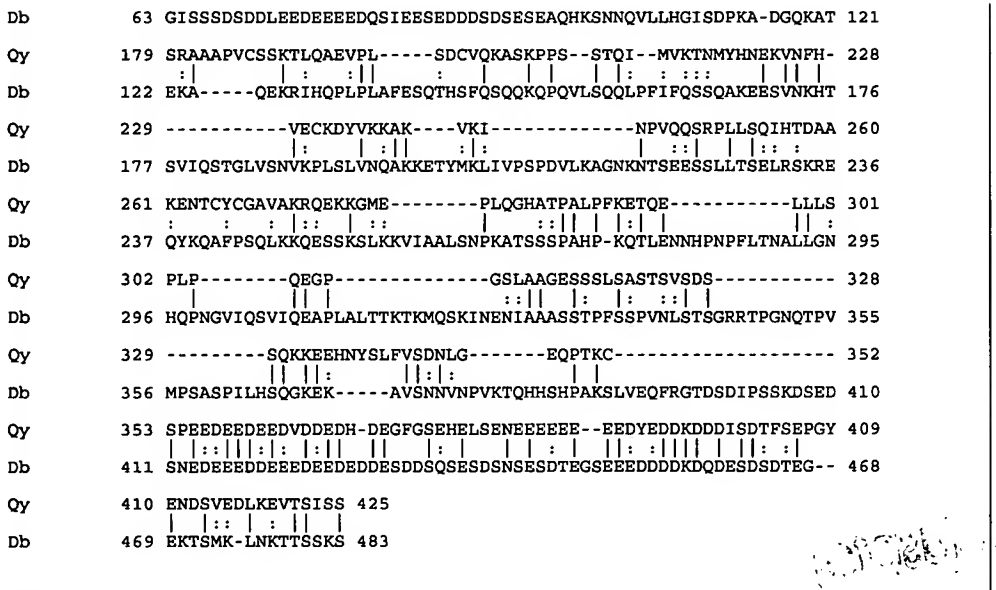
```
;  TITLE OF INVENTION:  Methods to Inhibit or Enhance the Binding of Viral DNA
;  TITLE OF INVENTION:  to Genomic Host DNA
;  FILE REFERENCE:  UM-03778
;  CURRENT APPLICATION NUMBER: US/09/410,399
;  CURRENT FILING DATE:  1999-10-01
;  NUMBER OF SEQ ID NOS:  6
;  SOFTWARE:  PatentIn Ver. 2.0
; SEQ ID NO 2
;   LENGTH: 1162
;   TYPE: PRT
;   ORGANISM: Kaposi's sarcoma-associated herpesvirus
US-09-410-399-2

  Query Match             8.0%;  Score 190;  DB 2;  Length 1162;
  Best Local Similarity   24.2%;  Pred. No. 1.4e-07;
  Matches   94;  Conservative   50;  Mismatches  137;  Indels  108;  Gaps   14;

Qy        124 SSPYQDEEVISKTPTLAQLNSEDSQSVSDSLYYPDSLFSVKQNPLPSSFPGKKITSRAAA 183
              |||      :|   |       |: | | :|    | | :: || ||  |:  :|
Db        111 SSPIPPSHPVSPGTT-------DTHSPSPALPPTQSPESSQRPPL-SSPTGRPDSSTPMR 162

Qy        184 PVCSSKTLQAEVPLSDCVQKASK------PPSSTQIMVKTNMYHNEKVNFHVECKDYVKK 237
              |  | :|     | :   :  ||       ||: : : |  :                 :
Db        163 PPPSQQTTPPHSPTTPPPEPPSKSSPDSLAPSTLRSLRKRRLSS--------------PQ 208

Qy        238 AKVKINPVQQSRPLL--------SQIHTDAAKENTCYCGAVAKRQEKKGMEPLQ----GH 285
              :||: || |:          ::   | |:  | |: :          |     |
Db        209 GPSTLNPICQSPPVSPPRCDFANRSVYPPWATESPIYVGSSSDGDTPPRQPPTSPISIGS 268

Qy        286 ATPALPF--KETQELLL--------SPLPQEGPGSLAAGESSSLSASTSVSDSSQKKEEH 335
              ::|:      :|  |:|        |   :|   :  |||    :    :|  ||  ::
Db        269 SSPSEGSWGDDTAMLVLLAEIAEEASKNEKECSENNQAGED---NGDNEISKESQVDKDD 325

Qy        336 NYSLFVSDNLGEQPTKCSPEEDEEDEE--------------------------DVDDED 368
              |      ||  ::  : : ||||||:|                          : |||:
Db        326 N------DNKDDEEEQETDEEDEEDDEEDDEEDDEEDDEEDDEEDDEEDDEEDDEEDDEE 379

Qy        369 HDEGFGSEHELSENEEEEEEEEDYEDDKDDDISDTFSEPGYENDSVEDLKE--------- 419
               ||    | :  |:|||:||||| ||| |:|  |        |:|  || ||
Db        380 DDEEDDEEEDEEEDEEEDEEEEDEEDDDDEDNEDE------EDDEEEDKKEDEEDGGDGN 433

Qy        420 -VTSISSRKRGKRRYFWEYSEQLTPSQQE 447
               || | :: :        :|  | |||
Db        434 KTLSIQSSQQQQE------PQQQEPQQQE 456


RESULT 5
US-09-894-273-2
; Sequence 2, Application US/09894273
; Patent No. 6756203
; GENERAL INFORMATION:
;  APPLICANT: Kieff, Elliott D.
;  APPLICANT:  Ballestas, Mary E.
;  APPLICANT:  Kaye, Kenneth M.
;  TITLE OF INVENTION: RHADINO VIRUS LANA ACTS IN TRANS ON A UNIT OF RHADINO
;  TITLE OF INVENTION:  VIRUS DNA TO MEDIATE EFFICIENT EPISOME PERSISTENCE
;  FILE REFERENCE: 16412-10001R
;  CURRENT APPLICATION NUMBER: US/09/894,273
;  CURRENT FILING DATE:  2001-06-28
;  PRIOR APPLICATION NUMBER: US 60/109,422
;  PRIOR FILING DATE: 1998-11-19
;  NUMBER OF SEQ ID NOS: 3
;  SOFTWARE: PatentIn Ver. 2.0
; SEQ ID NO 2
;   LENGTH: 1162
;   TYPE: PRT
;   ORGANISM: Kaposi's sarcoma-associated herpesvirus
US-09-894-273-2

  Query Match             8.0%;  Score 190;  DB 2;  Length 1162;
  Best Local Similarity   24.2%;  Pred. No. 1.4e-07;
  Matches   94;  Conservative   50;  Mismatches  137;  Indels  108;  Gaps   14;

Qy        124 SSPYQDEEVISKTPTLAQLNSEDSQSVSDSLYYPDSLFSVKQNPLPSSFPGKKITSRAAA 183
              |||      :|   |       |: | | :|    | | :: || ||  |:  :|
Db        111 SSPIPPSHPVSPGTT-------DTHSPSPALPPTQSPESSQRPPL-SSPTGRPDSSTPMR 162

Qy        184 PVCSSKTLQAEVPLSDCVQKASK------PPSSTQIMVKTNMYHNEKVNFHVECKDYVKK 237
              |  | :|     | :   :  ||       ||: : : |  :                 :
Db        163 PPPSQQTTPPHSPTTPPPEPPSKSSPDSLAPSTLRSLRKRRLSS--------------PQ 208

Qy        238 AKVKINPVQQSRPLL--------SQIHTDAAKENTCYCGAVAKRQEKKGMEPLQ----GH 285
              :||: || |:          ::   | |:  | |: :          |     |
Db        209 GPSTLNPICQSPPVSPPRCDFANRSVYPPWATESPIYVGSSSDGDTPPRQPPTSPISIGS 268

Qy        286 ATPALPF--KETQELLL--------SPLPQEGPGSLAAGESSSLSASTSVSDSSQKKEEH 335
              ::|:      :|  |:|        |   :|   :  |||    :    :|  ||  ::
Db        269 SSPSEGSWGDDTAMLVLLAEIAEEASKNEKECSENNQAGED---NGDNEISKESQVDKDD 325

Qy        336 NYSLFVSDNLGEQPTKCSPEEDEEDEE--------------------------DVDDED 368
              |      ||  ::  : : ||||||:|                          : |||:
Db        326 N------DNKDDEEEQETDEEDEEDDEEDDEEDDEEDDEEDDEEDDEEDDEEDDEEDDEE 379

Qy        369 HDEGFGSEHELSENEEEEEEEEDYEDDKDDDISDTFSEPGYENDSVEDLKE--------- 419
               ||    | :  |:|||:||||| ||| |:|  |        |:|  || ||
Db        380 DDEEDDEEEDEEEDEEEDEEEEDEEDDDDEDNEDE------EDDEEEDKKEDEEDGGDGN 433

Qy        420 -VTSISSRKRGKRRYFWEYSEQLTPSQQE 447
               || | :: :        :|  | |||
Db        434 KTLSIQSSQQQQE------PQQQEPQQQE 456


RESULT 6
US-09-538-092-957
; Sequence 957, Application US/09538092
; Patent No. 6753314
; GENERAL INFORMATION:
;  APPLICANT: Giot, Loic
;  APPLICANT:  Mansfield, Traci A.
;  TITLE OF INVENTION: Protein-Protein Complexes and Method of Using Same
;  FILE REFERENCE: 15966-542
;  CURRENT APPLICATION NUMBER: US/09/538,092
;  CURRENT FILING DATE:  2000-03-29
;  PRIOR APPLICATION NUMBER: 60/127,352
;  PRIOR FILING DATE: 1999-04-01
;  PRIOR APPLICATION NUMBER: 60/178,965
;  PRIOR FILING DATE: 2000-02-01
;  NUMBER OF SEQ ID NOS: 1387
;  SOFTWARE: CuraPatSeqFormatter Version 0.9
; SEQ ID NO 957
;   LENGTH: 706
;   TYPE: PRT
;   ORGANISM: Homo sapiens
;   FEATURE:
;   NAME/KEY: misc_feature
;   LOCATION: (0)...(0)
;   OTHER INFORMATION: Polypeptide Accession Number P19338
US-09-538-092-957

  Query Match             7.0%;  Score 165;  DB 2;  Length 706;
  Best Local Similarity   22.0%;  Pred. No. 1e-05;
  Matches   76;  Conservative   52;  Mismatches  123;  Indels   94;  Gaps   12;

Qy        129 DEEVISKTPTLAQLNSEDSQSVSDSLYYPDSLFSVKQNPLPSSFPGKKITSRAAAPVCSS 188
              | : ::  |   : :|||  :   |     ||  |  ::  :|    |||  : :|  |  |
Db         12 DPKKMAPPPKEVEEDSEDEEMSEDE--EDDS--SGEEVVIPQK-KGKKAAATSAKKVVVS 66

Qy        189 KTLQAEVPLSDCVQKASKPPSSTQIMVKTNMYHNEKVNFHVECKDYVKKAKVKINPVQQS 248
               |   :| ::   :||:  |                        |      ||  : | :
Db         67 PT--KKVAVATPAKKAAVTPGK--------------------KAAATPAKKTVTPAK-- 101

Qy        249 RPLLSQIHTDAAKENTCYCGAVAKRQEKKGMEPLQGHATPALPFKETQELLLSPLPQEGP 308
                                  |||  |  | |  | | |:       : :| :|
Db        102 --------------------AVTTPGKKGATP--GKALVATPGKKG-----AAIPAKGA 133

Qy        309 GSLAAGESSSLSASTSVSDSSQKKEEHNYSLFVSDNLGEQP--------TKCSPEEDEED 360
```

```
Db        134 KN---GKNAKKEDSDEEEDDDSEEDEEDDEDEDEDEDEIEPAAMKAAAAAPASEDEDDED 190

Qy        361 EEDVDDEDHDEGFGSEHE----------------------LSENEEEEEEEEDYEDDKDD 398

Db        191 DEDDEDDDDDEEDDSEEEAMETTPAKGKKAAKVVPVKAKNVAEDEDEEEDDEDEDDDDDE 250

Qy        399 DISDTFSEPGYENDSVED---LKEVTSISSRKRGKRRYFWEYSEQ 440

Db        251 DDEDDDDEDDEEEEEEEEEPVKEAPGKRKKEMAKQKAAPEAKKQ 295


RESULT 7
US-09-949-016-10040
; Sequence 10040, Application US/09949016
; Patent No. 6812339
; GENERAL INFORMATION:
;  APPLICANT: VENTER, J. Craig et al.
;  TITLE OF INVENTION: POLYMORPHISMS IN KNOWN GENES ASSOCIATED
;  TITLE OF INVENTION:  WITH HUMAN DISEASE, METHODS OF DETECTION AND USES THEREOF
;  FILE REFERENCE: CL001307
;  CURRENT APPLICATION NUMBER: US/09/949,016
;  CURRENT FILING DATE:  2000-04-14
;  PRIOR APPLICATION NUMBER: 60/241,755
;  PRIOR FILING DATE: 2000-10-20
;  PRIOR APPLICATION NUMBER: 60/237,768
;  PRIOR FILING DATE: 2000-10-03
;  PRIOR APPLICATION NUMBER: 60/231,498
;  PRIOR FILING DATE: 2000-09-08
;  NUMBER OF SEQ ID NOS: 207012
;  SOFTWARE: FastSEQ for Windows Version 4.0
; SEQ ID NO 10040
;   LENGTH: 747
;   TYPE: PRT
;   ORGANISM: Human
US-09-949-016-10040

  Query Match             7.0%;  Score 165;  DB 2;  Length 747;
  Best Local Similarity   22.0%;  Pred. No. 1.1e-05;
  Matches   76;  Conservative   52;  Mismatches  123;  Indels   94;  Gaps   12;

Qy        129 DEEVISKTPTLAQLNSEDSQSVSDSLYYPDSLFSVKQNPLPSSFPGKKITSRAAAPVCSS 188

Db         50 DPKKMAPPPKEVEEDSEDEEMSEDE--EDDS--SGEEVVIPQK-KGKKAAATSAKKVVVS 104

Qy        189 KTLQAEVPLSDCVQKASKPPSSTQIMVKTNMYHNEKVNFHVECKDYVKKAKVKINPVQQS 248

Db        105 PT--KKVAVATPAKKAAVTPGK--------------------KAAATPAKKTVTPAK-- 139

Qy        249 RPLLSQIHTDAAKENTCYCGAVAKRQEKKGMEPLQGHATPALPFKETQELLLSPLPQEGP 308

Db        140 ---------------------AVTTPGKKGATP--GKALVATPGKKG-----AAIPAKGA 171

Qy        309 GSLAAGESSSLSASTSVSDSSQKKEEHNYSLFVSDNLGEQP--------TKCSPEEDEED 360

Db        172 KN---GKNAKKEDSDEEEDDDSEEDEEDDEDEDEDEDEIEPAAMKAAAAAPASEDEDDED 228

Qy        361 EEDVDDEDHDEGFGSEHE----------------------LSENEEEEEEEEDYEDDKDD 398

Db        229 DEDDEDDDDDEEDDSEEEAMETTPAKGKKAAKVVPVKAKNVAEDEDEEEDDEDEDDDDDE 288

Qy        399 DISDTFSEPGYENDSVED---LKEVTSISSRKRGKRRYFWEYSEQ 440

Db        289 DDEDDDDEDDEEEEEEEEEPVKEAPGKRKKEMAKQKAAPEAKKQ 333


RESULT 8
US-09-248-796A-19045
; Sequence 19045, Application US/09248796A
; Patent No. 6747137
; GENERAL INFORMATION:
;  APPLICANT: Keith Weinstock et al
;  TITLE OF INVENTION: NUCLEIC ACID AND AMINO ACID SEQUENCES RELATING TO CANDIDA ALBICAN
;  TITLE OF INVENTION:  FOR DIAGNOSTICS AND THERAPEUTICS
;  FILE REFERENCE: 107196.132
;  CURRENT APPLICATION NUMBER: US/09/248,796A
;  CURRENT FILING DATE:  1999-02-12
;  PRIOR APPLICATION NUMBER: US 60/074,725
;  PRIOR FILING DATE: 1998-02-13
;  PRIOR APPLICATION NUMBER: US 60/096,409
;  PRIOR FILING DATE: 1998-08-13
;  NUMBER OF SEQ ID NOS: 28208
; SEQ ID NO 19045
;   LENGTH: 709
;   TYPE: PRT
;   ORGANISM: Candida albicans
US-09-248-796A-19045

  Query Match             6.8%;  Score 161.5;  DB 2;  Length 709;
  Best Local Similarity   19.0%;  Pred. No. 2.1e-05;
  Matches  115;  Conservative   79;  Mismatches  196;  Indels  215;  Gaps   24;

Qy         26 TLMSTDLLANSSDPDF---------MYELDREMNYQQ------------NPRDNFLSLED 64

Db        105 TESSTDVISNTDYNDIDNSTTAETPFTELEKDTVSESLDTTTAPKTSTTNSDDDVAQLND 164

Qy         65 CKDIENLESFTDVLDNEGALTSNWEQWDTYCEDLTKYTKLTSCDIWGTKEVDYLGLDDFS 124

Db        165 KKSIVSFGSFDEQID---------EKTEESLRELKSLNRLS----------DKLEIKNSD 205

Qy        125 SP----------YQDEEVISKTPTLAQLNSEDSQSVS--------------------DSL 154

Db        206 TPNKLTNAAINDVQEKTKPEKVSTL--LNDDDEQKVKTEVNVELKDKTEEEKKSEEINKL 263

Qy        155 YYPDSLFSVKQNP---LPSSFPGKKITSRAAAPVCSSKTLQAEVPLSDCV---------- 201

Db        264 TPPDVVID-KDEPVKTLDESLGETKEIKKDTAVKPSTTT--KEVSKNDSIVEPKTIKFHT 320

Qy        202 -------------QKASKPPSSTQIMVKTNMYHNEKVNFHV-ECKDYVKKAKVKINPVQQ 247

Db        321 EVKGETKSDSTTEPKVKETPKDTKEEIKTDAIVEPKVAKPIDESKDDIKKDSVSAEP-KV 379

Qy        248 SRPLLSQIHTDAAKENTCYC---GAVAKR---------QEKKGMEPLQGHATPAL----- 290

Db        380 VEPL--DEHKEDKTETAKICTDTSTIAEQPPVKTEIVDSKKKEPEPLADDKTKASEDIKD 437

Qy        291 ------------------------------------------PFKETQEL------LL 300

Db        438 VNKIEKKSGNKETMVDTKLETTLDKEVDTDSNIQEKKELFQIKEPVKDEKQIKEDLVKVE 497

Qy        301 SPLPQEGPGSLAAGESS----SLSASTSVSDSSQKKEEHNYSLFVSDNL----------- 345

Db        498 TPSKDEGPAKVEVPSKDETPFKIQNIDSTTDNGKAKDEKDNVSKVDSNVDTKVVTKDDQD 557

Qy        346 GEQPTKCSPEEDEEDEE----------DVDDEDHDEGFGSEHELSENEEEEEEEED-- 391

Db        558 GIKKVEEKPVEKKDDDEGKKDNEKVTVHKDNEEEDKDKGVNEAKDKVQKDDDEEEDIDKD 617

Qy        392 ------YEDDKDDDISDTFSEPG------YENDSVEDLKEVTSISSRKRGKRRYFWEYSE 439

Db        618 LVIVNKEEEDKDKDVKDDNKDVNKAVDNDKENENEED--DYNTDNENNTLKTPYDFEASP 675

Qy        440 QLTPS 444

Db        676 SPSPS 680


RESULT 9
US-09-047-026A-4
; Sequence 4, Application US/09047026A
; Patent No. 5989897
;  GENERAL INFORMATION:
;    APPLICANT:  Pillus, Lorraine
;    APPLICANT:  Clarke, Astrid
;    APPLICANT:  Lowell, Joanna
;    APPLICANT:  Jacobson, Sandra
```

```
;    APPLICANT:  Reifsnyder, Cheryl
;    TITLE OF INVENTION:  Yeast Silencing Genes, Proteins and
;    TITLE OF INVENTION:  Methods
;    NUMBER OF SEQUENCES:  25
;    CORRESPONDENCE ADDRESS:
;      ADDRESSEE:  Greenlee, Winner and Sullivan, P.C.
;      STREET:  5370 Manhattan Circle, Suite 201
;      CITY:  Boulder
;      STATE:  Colorado
;      COUNTRY:  US
;      ZIP:  80303
;    COMPUTER READABLE FORM:
;      MEDIUM TYPE:  Floppy disk
;      COMPUTER:  IBM PC compatible
;      OPERATING SYSTEM:  PC-DOS/MS-DOS
;      SOFTWARE:  PatentIn Release #1.0, Version #1.30
;    CURRENT APPLICATION DATA:
;      APPLICATION NUMBER:  US/09/047,026A
;      FILING DATE:  24-MAR-1998
;      CLASSIFICATION:
;    PRIOR APPLICATION DATA:
;      APPLICATION NUMBER:  US 60/042,375
;      FILING DATE:  24-MAR-1997
;    ATTORNEY/AGENT INFORMATION:
;      NAME:  Ferber, Donna M.
;      REGISTRATION NUMBER:  33,878
;      REFERENCE/DOCKET NUMBER:  1-97
;    TELECOMMUNICATION INFORMATION:
;      TELEPHONE:  (303) 499-8080
;      TELEFAX:  (303) 499-8089
;  INFORMATION FOR SEQ ID NO:  4:
;    SEQUENCE CHARACTERISTICS:
;      LENGTH:  831 amino acids
;      TYPE:  amino acid
;      TOPOLOGY:  linear
;    MOLECULE TYPE:  protein
US-09-047-026A-4

  Query Match             6.8%;  Score 161;  DB 1;  Length 831;
  Best Local Similarity   20.6%;  Pred. No. 3e-05;
  Matches   86;  Conservative   62;  Mismatches  116;  Indels  154;  Gaps   20;

Qy          86 SNWEQWDTY----CEDLTKYTKLTSCDIWGTKEVDYLGLDDFSSPYQDEEVISKTPTLAQ 141
               || :  ||:       |: | | :   |:          ||:      |:      || :|:
Db         486 SNNKNEDTFQQVSLNDIAKLTGMIPTDV-------VFGLEQLQVLYR-----HKTRSLSS 533

Qy         142 LN------SEDSQSVSDSLY-------YP----DSL----------FSVK--QNPLPSSF 172
               |:        || :  :::|       ||    | |          | :    |  |::
Db         534 LDDFNYIIKIDSWNRIENIYKTWSSKNYPRVKYDKLLWEPIILGPSFGINGMMNLEPTAL 593

Qy         173 PGKKITSRAAAPVCSSKT---------------------------------LQAEVPLS 198
                 : :|:   ||| |: |                                 :: ||| :
Db         594 ADEALTNETMAPVISNNTHIENYNNSRAHNKRRRRRRRSSEHKTSKLHVNNIIEPEVPAT 653

Qy         199 DCVQKASKPPSSTQIMVKTNMYHNEKVNFHVECKDYVKKAKVKINPVQQSRPLLSQIHTD 258
               |  :      | |: |      :|::: :  |                    : :|  ||
Db         654 DFFEDTVS--SLTEYMCDYKNTNNDRLIYQAE-----------------KRVLESIH-- 691

Qy         259 AAKENTCYCGAVAKRQEKKGMEPLQGHATPALPFK-ETQ-ELLLSPLPQEGPGSLAAGES 316
                           ::||:         |    |  ||  ||  :    | |     |
Db         692 ---------------DRKGI--------PRSKFSTETHWELCFTIKNSETP----LGNH 723

Qy         317 SSLSASTSVSDSSQKKEEHNY--SLFVSDNLGEQPTKCSPEEDEEDEEDVDDEDHDEGFG 374
               ::    | :|    | : |::    |:| :|  |        :|||::  :||:  ||
Db         724 AARRNDTGISSLEQDEVENDVDTELYVGENAKE-------DEDEDEDFTLDDDIEDEQI- 775

Qy         375 SEHELSENEEEEEEEEDYEDDKDDDISDTFSEPGYENDSVEDLKEVTSISSRKRGKRR 432
                     :|| :|||: ||:| |||       | |      :|  :: |   ::| ::|
Db         776 -------SEENDEEEDTYEEDSDDD------EDGKRKGQEQDENDIESHIRKERVRKR 820


RESULT 10
US-09-418-710-21
```

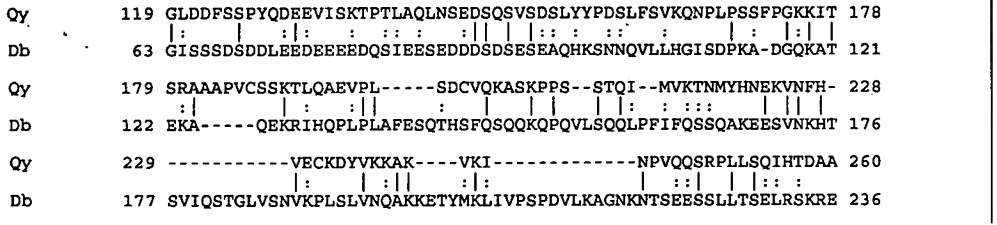
```
; Sequence 21, Application US/09418710
; Patent No. 6596482
; GENERAL INFORMATION:
;  APPLICANT:  Jones, Michael H.
;  TITLE OF INVENTION:  TRANSCRIPTIONAL REGULATOR
;  FILE REFERENCE:  06501-042001
;  CURRENT APPLICATION NUMBER: US/09/418,710
;  CURRENT FILING DATE:  1999-10-15
;  PRIOR APPLICATION NUMBER:  PCT/JP98/01783
;  PRIOR FILING DATE:  1998-04-17
;  PRIOR APPLICATION NUMBER:  JP 9/310027
;  PRIOR FILING DATE:  1997-10-24
;  PRIOR APPLICATION NUMBER:  JP 9/116570
;  PRIOR FILING DATE:  1997-04-18
;  NUMBER OF SEQ ID NOS:  73
;  SOFTWARE:  FastSEQ for Windows Version 4.0
; SEQ ID NO 21
;   LENGTH: 1972
;   TYPE: PRT
;   ORGANISM: Homo sapiens
US-09-418-710-21

  Query Match             6.8%;  Score 160;  DB 2;  Length 1972;
  Best Local Similarity   21.5%;  Pred. No. 0.00014;
  Matches   94;  Conservative   68;  Mismatches  130;  Indels  146;  Gaps   20;

Qy         119 GLDDFSSPYQDEEVISKTPTLAQLNSEDSQSVSDSLYYPDSLFSVKQNPLPSSFPGKKIT 178
                 |:     |    :|:    :  :: :   :|| | |:: :  ::    :      | :  |:| |
Db          63 GISSSDSDDLEEDEEEEDQSIEESEDDDSDSESEAQHKSNNQVLLHGISDPKA-DGQKAT 121

Qy         179 SRAAAPVCSSKTLQAEVPL-----SDCVQKASKPPS--STQI--MVKTNMYHNEKVNFH- 228
                  :|        | :   :||     :    |   | |   | |:  : :::     | || |
Db         122 EKA-----QEKRIHQPLPLAFESQTHSFQSQQKQPQVLSQQLPFIFQSSQAKEESVNKHT 176

Qy         229 -----------VECKDYVKKAK----VKI-------------NPVQQSRPLLSQIHTDAA 260
                         |:    | :||    :|:               |  ::|  | |:: :
Db         177 SVIQSTGLVSNVKPLSLVNQAKKETYMKLIVPSPDVLKAGNKNTSEESSLLTSELRSKRE 236

Qy         261 KENTCYCGAVAKRQEKKGME--------PLQGHATPALPFKETQE-----------LLLS 301
                  :   :  : |::  | ::         |     ::|| | |:| |           || :
Db         237 QYKQAFPSQLKKQESSKSLKKVIAALSNPKATSSSPAHP-KQTLENNHPNPFLTNALLGN 295

Qy         302 PLP--------QEGP-------------GSLAAGESSSLSASTSVSDS----------- 328
                 |        || |                ::||  |:  |:  ::| |
Db         296 HQPNGVIQSVIQEAPLALTTKTKMQSKINENIAAASSTPFSSPVNLSTSGRRTPGNQTPV 355

Qy         329 ---------SQKKEEHNYSLFVSDNLG-------EQPTKC-------------------- 352
                       || ||:       ||:|:           | |
Db         356 MPSASPILHSQGKEK-----AVSNNVNPVKTQHHSHPAKSLVEQFRGTDSDIPSSKDSED 410

Qy         353 ---SPEEDEEDEEDVDDEDHDEGFGSEHELSENEEEEEE--EEDYEDDKDDDISDTFSEP 407
                    |||:|:|:: |||| ||   |: |   | | : |  ||: :|||| | ||: :|
Db         411 SNEDEEEDDEEEDEEDDED-DESDDSQSESDSNSESDTEGSEEEDDDDKDQDESDSDTEG 469

Qy         408 GYENDSVEDLKEVTSISS 425
                 |  |::  |  :|: |
Db         470 --EKTSMKLNKTTSSVKS 485


RESULT 11
US-09-839-479-21
; Sequence 21, Application US/09839479
; Patent No. 6727222
; GENERAL INFORMATION:
;  APPLICANT: Jones, Michael H.
;  TITLE OF INVENTION: TRANSCRIPTIONAL REGULATOR
;  FILE REFERENCE: 06501-042002
;  CURRENT APPLICATION NUMBER: US/09/839,479
;  CURRENT FILING DATE:  2001-04-20
;  PRIOR APPLICATION NUMBER: US 09/418,710
;  PRIOR FILING DATE: 1999-10-15
;  PRIOR APPLICATION NUMBER: PCT/JP98/01783
;  PRIOR FILING DATE: 1998-04-17
```

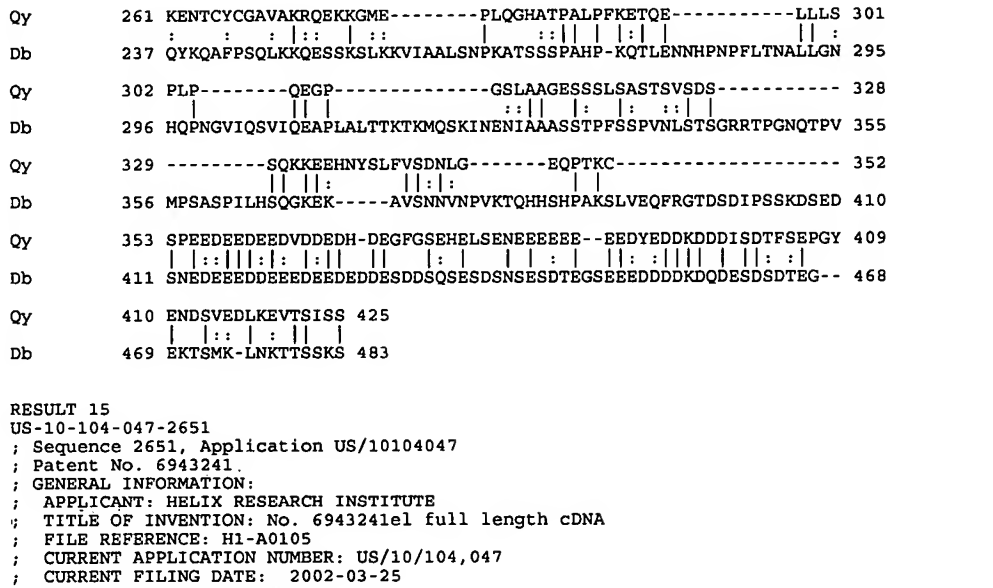
```
;  PRIOR APPLICATION NUMBER: JP 9/310027
;  PRIOR FILING DATE: 1997-10-24
;  PRIOR APPLICATION NUMBER: JP 9/116570
;  PRIOR FILING DATE: 1997-04-18
;  NUMBER OF SEQ ID NOS: 72
;  SOFTWARE: FastSEQ for Windows Version 4.0
; SEQ ID NO 21
;   LENGTH: 1972
;   TYPE: PRT
;   ORGANISM: Homo sapiens
US-09-839-479-21

  Query Match             6.8%;  Score 160;  DB 2;  Length 1972;
  Best Local Similarity   21.5%;  Pred. No. 0.00014;
  Matches   94;  Conservative   68;  Mismatches  130;  Indels  146;  Gaps   20;

Qy        119 GLDDFSSPYQDEEVISKTPTLAQLNSEDSQSVSDSLYYPDSLFSVKQNPLPSSFPGKKIT 178
              |:    |   :|:   :  :: :   :|| | |:: :  ::   :      | :  |:| |
Db         63 GISSSDSDDLEEDEEEEDQSIEESEDDDSDSESEAQHKSNNQVLLHGISDPKA-DGQKAT 121

Qy        179 SRAAAPVCSSKTLQAEVPL-----SDCVQKASKPPS--STQI--MVKTNMYHNEKVNFH- 228
              :|       | :   :||     :   |  | |   | |:  : :::    | || |
Db        122 EKA-----QEKRIHQPLPLAFESQTHSFQSQQKQPQVLSQQLPFIFQSSQAKEESVNKHT 176

Qy        229 -----------VECKDYVKKAK----VKI------------NPVQQSRPLLSQIHTDAA 260
                        |:    | :||    :|:             |  ::|  | |:: :
Db        177 SVIQSTGLVSNVKPLSLVNQAKKETYMKLIVPSPDVLKAGNKNTSEESSLLTSELRSKRE 236

Qy        261 KENTCYCGAVAKRQEKKGME--------PLQGHATPALPFKETQE-----------LLLS 301
              :   :   : |::  | ::         |     ::|| | |:| |           || :
Db        237 QYKQAFPSQLKKQESSKSLKKVIAALSNPKATSSSPAHP-KQTLENNHPNPFLTNALLGN 295

Qy        302 PLP--------QEGP-------------GSLAAGESSSLSASTSVSDS----------- 328
              |         || |               ::||  |:  |:  ::| |
Db        296 HQPNGVIQSVIQEAPLALTTKTKMQSKINENIAAASSTPFSSPVNLSTSGRRTPGNQTPV 355

Qy        329 ---------SQKKEEHNYSLFVSDNLG-------EQPTKC-------------------- 352
                      || ||:      ||:|:          | |
Db        356 MPSASPILHSQGKEK-----AVSNNVNPVKTQHHSHPAKSLVEQFRGTDSDIPSSKDSED 410

Qy        353 ---SPEEDEEDEEDVDDEDHDEGFGSEHELSENEEEEEE--EEDYEDDKDDDISDTFSEP 407
                  |||:|:|:: |||| ||   |: |   | | : |  ||: :|||| | ||: :|
Db        411 SNEDEEEDDEEEDEEDDED-DESDDSQSESDSNSESDTEGSEEEDDDDKDQDESDSDTEG 469

Qy        408 GYENDSVEDLKEVTSISS 425
                |  |::  |  :|: |
Db        470 --EKTSMKLNKTTSSVKS 485


RESULT 12
US-09-248-796A-16666
; Sequence 16666, Application US/09248796A
; Patent No. 6747137
; GENERAL INFORMATION:
;  APPLICANT: Keith Weinstock et al
;  TITLE OF INVENTION: NUCLEIC ACID AND AMINO ACID SEQUENCES RELATING TO CANDIDA ALBICAN
;  TITLE OF INVENTION:  FOR DIAGNOSTICS AND THERAPEUTICS
;  FILE REFERENCE: 107196.132
;  CURRENT APPLICATION NUMBER: US/09/248,796A
;  CURRENT FILING DATE:  1999-02-12
;  PRIOR APPLICATION NUMBER: US 60/074,725
;  PRIOR FILING DATE: 1998-02-13
;  PRIOR APPLICATION NUMBER: US 60/096,409
;  PRIOR FILING DATE: 1998-08-13
;  NUMBER OF SEQ ID NOS: 28208
; SEQ ID NO 16666
;   LENGTH: 738
;   TYPE: PRT
;   ORGANISM: Candida albicans
;   FEATURE:
;   NAME/KEY: UNSURE
;   LOCATION: (154)
;   OTHER INFORMATION: Identity of amino acid sequences at the above locations are unkno
US-09-248-796A-16666

  Query Match             6.7%;  Score 157.5;  DB 2;  Length 738;
  Best Local Similarity   20.3%;  Pred. No. 5.1e-05;
  Matches  102;  Conservative   82;  Mismatches  163;  Indels  155;  Gaps   24;

Qy         11 PPFGDAFRS--HTFSEQTLMSTDL--LANSSDPDFMYELDREMNYQQNPRDNFLSLEDCK 66
              |||  :  :  ||:| :   |: :  | ::: |           :| |  |: :: |  :
Db         12 PPFHHSTHTHIHTYSYKNTYSSIMTSLHDTTTP------TTPAPHQHNDSDSDIN-EMSR 64

Qy         67 DIENLESFTDVLDNEGALTSNWE----------QWDTYCEDLTKYTKLTSCDIWGTKEV 115
              |:: :      | |   ||| :           :   :  : |||          ||
Db         65 DLDEIG-----LLNLNNLTSYFSTDRFFSQMPTFEMKNFPLEFTKYV---------TKYK 110

Qy        116 DYLGLDDFSSPYQDE--EVISKTPTLAQLNSEDSQSVSDSLYYPDS-LFSVKQNPLPSSF 172
              :|          ||::  :       | |:  | |:  | | |   || :       | :
Db        111 NY---------YQNQYKKYNKSFGTTAKTESTDTDEVKDILSEDDSNILDSIXTVLSTRI 161

Qy        173 PGKKITSRAAAPVCSSKTLQAEVPLSDCVQKASKPPSSTQIMVKTNMY---HNEKVNFHV 229
              |  :       |  :  :    ||:| : :          | : |: :    |::    :
Db        162 QMDKFLNDLKLGVNITTAIDRLQPLTDVINET--------IFLPTDDFSDIENDENELDI 213

Qy        230 ECK---------DYVKKAKVKINPVQQSRPLLSQIHTDAAKENTCYCGAVAKRQEKKGME 280
              | :         : | |::  || :   :| : ::  :  : |         ||:: | :
Db        214 EIEEITEPSGIGEIVNKSQPLINEI--VKPRVGEMRNEIKRNNR----RKRKRKQYKNIY 267

Qy        281 PLQGHATPALPFKETQELLLSPLPQEGPGSLAAGESSSLSASTSVSD-------SSQKKE 333
                    |      :| ::            |   | | |:   |: |        |: |:
Db        268 SNNDSVT-----SQTSDI----------DGLTIYEVSELTELNSLDDFVKQDLLRSKIKK 312

Qy        334 EHNYSL-------------------FVSDNLGE------QP-------TKCSPEEDEEDE 361
               || |:                 :| : |          ||         |   |:||| :
Db        313 IHNLSISQDSKNKLVTKLMMGNYYKYVDEKLSNNDKQLLQPLRNQKLIIKEEKEKDEEKD 372

Qy        362 EDVDDEDHD----------------------EGFGSEHELSENEEEEEEEEDYEDDKDD 398
               | :||: |                     | | || | | :|||::||:| |||::|
Db        373 NDEEDEEDDFDEELPSRLSLGPVDEAIIENDGENFESESEGSNDEEEDDEEDDEEDDEED 432

Qy        399 D---ISDTFSEPGYENDSVEDL 417
                |   :::   :| | :||:  :
Db        433 DEVMLTEQDLQPTY-HDSLHTI 453


RESULT 13
US-09-418-710-72
; Sequence 72, Application US/09418710
; Patent No. 6596482
; GENERAL INFORMATION:
;  APPLICANT:  Jones, Michael H.
;  TITLE OF INVENTION:  TRANSCRIPTIONAL REGULATOR
;  FILE REFERENCE:  06501-042001
;  CURRENT APPLICATION NUMBER: US/09/418,710
;  CURRENT FILING DATE:  1999-10-15
;  PRIOR APPLICATION NUMBER:  PCT/JP98/01783
;  PRIOR FILING DATE:  1998-04-17
;  PRIOR APPLICATION NUMBER:  JP 9/310027
;  PRIOR FILING DATE:  1997-10-24
;  PRIOR APPLICATION NUMBER:  JP 9/116570
;  PRIOR FILING DATE:  1997-04-18
;  NUMBER OF SEQ ID NOS:  73
;  SOFTWARE:  FastSEQ for Windows Version 4.0
; SEQ ID NO 72
;   LENGTH: 1969
;   TYPE: PRT
;   ORGANISM: Homo sapiens
US-09-418-710-72

  Query Match             6.6%;  Score 157;  DB 2;  Length 1969;
  Best Local Similarity   21.8%;  Pred. No. 0.00025;
  Matches   95;  Conservative   68;  Mismatches  129;  Indels  144;  Gaps   21;

Qy        119 GLDDFSSPYQDEEVISKTPTLAQLNSEDSQSVSDSLYYPDSLFSVKQNPLPSSFPGKKIT 178
              |:    |   :|:   :  :: :   :|| | |:: :  ::   :      | :  |:| |
```



RESULT 14
US-09-839-479-71
; Sequence 71, Application US/09839479
; Patent No. 6727222
; GENERAL INFORMATION:
; APPLICANT: Jones, Michael H.
; TITLE OF INVENTION: TRANSCRIPTIONAL REGULATOR
; FILE REFERENCE: 06501-042002
; CURRENT APPLICATION NUMBER: US/09/839,479
; CURRENT FILING DATE: 2001-04-20
; PRIOR APPLICATION NUMBER: US 09/418,710
; PRIOR FILING DATE: 1999-10-15
; PRIOR APPLICATION NUMBER: PCT/JP98/01783
; PRIOR FILING DATE: 1998-04-17
; PRIOR APPLICATION NUMBER: JP 9/310027
; PRIOR FILING DATE: 1997-10-24
; PRIOR APPLICATION NUMBER: JP 9/116570
; PRIOR FILING DATE: 1997-04-18
; NUMBER OF SEQ ID NOS: 72
; SOFTWARE: FastSEQ for Windows Version 4.0
; SEQ ID NO 71
; LENGTH: 1969
; TYPE: PRT
; ORGANISM: Homo sapiens
US-09-839-479-71

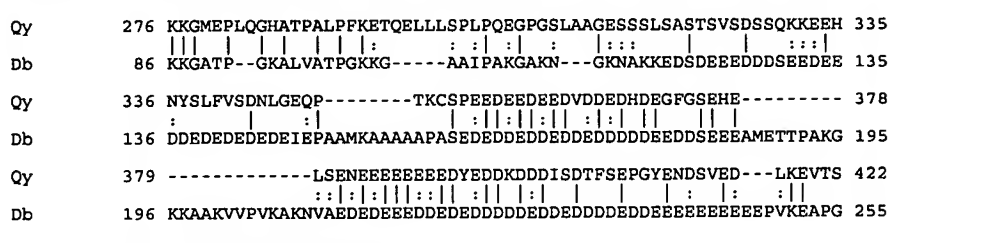
Query Match 6.6%; Score 157; DB 2; Length 1969;
Best Local Similarity 21.8%; Pred. No. 0.00025;
Matches 95; Conservative 68; Mismatches 129; Indels 144; Gaps 21;





RESULT 15
US-10-104-047-2651
; Sequence 2651, Application US/10104047
; Patent No. 6943241
; GENERAL INFORMATION:
; APPLICANT: HELIX RESEARCH INSTITUTE
; TITLE OF INVENTION: No. 6943241el full length cDNA
; FILE REFERENCE: H1-A0105
; CURRENT APPLICATION NUMBER: US/10/104,047
; CURRENT FILING DATE: 2002-03-25
; PRIOR APPLICATION NUMBER:
; PRIOR FILING DATE:
; NUMBER OF SEQ ID NOS: 4096
; SOFTWARE: PatentIn Ver. 2.1
; SEQ ID NO 2651
; LENGTH: 687
; TYPE: PRT
; ORGANISM: Homo sapiens
US-10-104-047-2651

Query Match 6.6%; Score 155.5; DB 2; Length 687;
Best Local Similarity 25.8%; Pred. No. 6.8e-05;
Matches 51; Conservative 34; Mismatches 70; Indels 43; Gaps 6;



Search completed: February 28, 2006, 08:56:07
Job time : 40.7997 secs

GenCore version 5.1.7

OM protein - protein search, using sw model

Run on: February 28, 2006, 08:54:36 ; Search time 130.62 Seconds
(without alignments)
1433.071 Million cell updates/sec

Title: US-10-717-665A-44_COPY_1_448
Perfect score: 2364
Sequence: 1 MPQPSVSGMDPPFGDAFRSH.....GKRRYFWEYSEQLTPSQQER 448

Scoring table: BLOSUM62
Gapop 10.0 , Gapext 0.5

Searched: 1867569 seqs, 417829326 residues

Total number of hits satisfying chosen parameters: 1867569

Minimum DB seq length: 0
Maximum DB seq length: 2000000000

Post-processing: Minimum Match 0%
Maximum Match 100%
Listing first 45 summaries

Database : Published_Applications_AA_Main:*
1: /cgn2_6/ptodata/1/pubpaa/US07_PUBCOMB.pep:*
2: /cgn2_6/ptodata/1/pubpaa/US08_PUBCOMB.pep:*
3: /cgn2_6/ptodata/1/pubpaa/US09_PUBCOMB.pep:*
4: /cgn2_6/ptodata/1/pubpaa/US10A_PUBCOMB.pep:*
5: /cgn2_6/ptodata/1/pubpaa/US10B_PUBCOMB.pep:*
6: /cgn2_6/ptodata/1/pubpaa/US11_PUBCOMB.pep:*

Pred. No. is the number of results predicted by chance to have a score greater than or equal to the score of the result being printed, and is derived by analysis of the total score distribution.

SUMMARIES

Result No.	Score	% Query Match	Length	DB	ID	Description
1	2364	100.0	639	5	US-10-717-665-44	Sequence 44, Appl
2	191.5	8.1	1003	5	US-10-732-923-3319	Sequence 3319, Ap
3	190	8.0	1162	3	US-09-894-273-2	Sequence 2, Appli
4	190	8.0	1162	4	US-10-294-804-2	Sequence 2, Appli
5	190	8.0	1162	5	US-10-194-046-2	Sequence 2, Appli
6	169.5	7.2	735	5	US-10-483-505-3	Sequence 3, Appli
7	165	7.0	707	3	US-09-825-886-22	Sequence 22, Appl
8	165	7.0	707	4	US-10-384-569-3	Sequence 3, Appli
9	165	7.0	707	4	US-10-400-083-19	Sequence 19, Appl
10	165	7.0	707	5	US-10-491-545A-20	Sequence 20, Appl
11	162	6.9	167	5	US-10-450-763-41629	Sequence 41629, A
12	160.5	6.8	382	5	US-10-450-763-34710	Sequence 34710, A
13	160.5	6.8	1974	4	US-10-369-493-5455	Sequence 5455, Ap
14	160	6.8	1972	3	US-09-839-479-21	Sequence 21, Appl
15	160	6.8	1972	4	US-10-376-537-21	Sequence 21, Appl
16	160	6.8	1972	4	US-10-702-148-21	Sequence 21, Appl
17	160	6.8	1972	5	US-10-489-740-176	Sequence 176, App
18	160	6.8	1972	5	US-10-756-149-5621	Sequence 5621, Ap
19	159.5	6.7	1128	5	US-10-732-923-3317	Sequence 3317, Ap
20	159	6.7	1038	6	US-11-097-143-37245	Sequence 37245, A
21	158	6.7	264	5	US-10-450-763-35100	Sequence 35100, A
22	157.5	6.7	647	5	US-10-691-672A-3	Sequence 3, Appli
23	157	6.6	146	5	US-10-450-763-35417	Sequence 35417, A
24	157	6.6	1969	3	US-09-839-479-71	Sequence 71, Appl
25	157	6.6	1969	4	US-10-376-537-72	Sequence 72, Appl
26	157	6.6	1969	4	US-10-702-148-71	Sequence 71, Appl
27	156	6.6	1251	4	US-10-112-944-881	Sequence 881, App
28	155.5	6.6	687	4	US-10-104-047-2651	Sequence 2651, Ap
29	155.5	6.6	687	6	US-11-013-684-7	Sequence 7, Appli
30	155.5	6.6	903	4	US-10-408-765A-1332	Sequence 1332, Ap
31	153.5	6.5	1214	5	US-10-450-763-52482	Sequence 52482, A
32	153	6.5	2724	5	US-10-487-593-1	Sequence 1, Appli
33	152.5	6.5	740	4	US-10-408-765A-1280	Sequence 1280, Ap
34	152.5	6.5	901	4	US-10-408-765A-1259	Sequence 1259, Ap
35	152.5	6.5	2400	5	US-10-450-763-50637	Sequence 50637, A
36	151	6.4	983	4	US-10-112-944-435	Sequence 435, App
37	151	6.4	1021	4	US-10-408-765A-313	Sequence 313, App
38	150.5	6.4	293	5	US-10-450-763-46978	Sequence 46978, A
39	150.5	6.4	681	6	US-11-097-143-30126	Sequence 30126, A
40	150.5	6.4	714	3	US-09-978-242-3	Sequence 3, Appli
41	150.5	6.4	740	4	US-10-012-697-1513	Sequence 1513, Ap
42	150.5	6.4	740	5	US-10-779-543-23513	Sequence 23513, A
43	149.5	6.3	376	4	US-10-156-761-9888	Sequence 9888, Ap
44	148.5	6.3	814	6	US-11-097-143-31914	Sequence 31914, A
45	148	6.3	407	5	US-10-450-763-34334	Sequence 34334, A

ALIGNMENTS

RESULT 1
US-10-717-665-44
; Sequence 44, Application US/10717665
; Publication No. US20050106579A1
; GENERAL INFORMATION:
; APPLICANT: OriGene Technologies, Inc
; TITLE OF INVENTION: Regulated Angiogenesis Genes and Polypeptides
; FILE REFERENCE: 1U 103 R1
; CURRENT APPLICATION NUMBER: US/10/717,665
; CURRENT FILING DATE: 2003-11-21
; PRIOR APPLICATION NUMBER: US/10/164,595
; PRIOR FILING DATE: 2002-06-10
; NUMBER OF SEQ ID NOS: 80
; SOFTWARE: PatentIn version 3.1
; SEQ ID NO 44
; LENGTH: 639
; TYPE: PRT
; ORGANISM: Homo sapiens
US-10-717-665-44

Query Match 100.0%; Score 2364; DB 5; Length 639;
Best Local Similarity 100.0%; Pred. No. 9.4e-153;
Matches 448; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

```
Qy          1 MPQPSVSGMDPPFGDAFRSHTFSEQTLMSTDLLANSSDPDFMYELDREMNYQQNPRDNFL 60
              ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db          1 MPQPSVSGMDPPFGDAFRSHTFSEQTLMSTDLLANSSDPDFMYELDREMNYQQNPRDNFL 60

Qy         61 SLEDCKDIENLESFTDVLDNEGALTSNWEQWDTYCEDLTKYTKLTSCDIWGTKEVDYLGL 120
              ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db         61 SLEDCKDIENLESFTDVLDNEGALTSNWEQWDTYCEDLTKYTKLTSCDIWGTKEVDYLGL 120

Qy        121 DDFSSPYQDEEVISKTPTLAQLNSEDSQSVSDSLYYPDSLFSVKQNPLPSSFPGKKITSR 180
              ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db        121 DDFSSPYQDEEVISKTPTLAQLNSEDSQSVSDSLYYPDSLFSVKQNPLPSSFPGKKITSR 180

Qy        181 AAAPVCSSKTLQAEVPLSDCVQKASKPPSSTQIMVKTNMYHNEKVNFHVECKDYVKKAKV 240
              ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db        181 AAAPVCSSKTLQAEVPLSDCVQKASKPPSSTQIMVKTNMYHNEKVNFHVECKDYVKKAKV 240

Qy        241 KINPVQQSRPLLSQIHTDAAKENTCYCGAVAKRQEKKGMEPLQGHATPALPFKETQELLL 300
              ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db        241 KINPVQQSRPLLSQIHTDAAKENTCYCGAVAKRQEKKGMEPLQGHATPALPFKETQELLL 300

Qy        301 SPLPQEGPGSLAAGESSSLSASTSVSDSSQKKEEHNYSLFVSDNLGEQPTKCSPEEDEED 360
              ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db        301 SPLPQEGPGSLAAGESSSLSASTSVSDSSQKKEEHNYSLFVSDNLGEQPTKCSPEEDEED 360

Qy        361 EEDVDDEDHDEGFGSEHELSENEEEEEEEEDYEDDKDDDISDTFSEPGYENDSVEDLKEV 420
              ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
```

```
Db        361 EEDVDDEDHDEGFGSEHELSENEEEEEEEEEDYEDDKDDDISDTFSEPGYENDSVEDLKEV 420

Qy        421 TSISSRKRGKRRYFWEYSEQLTPSQQER 448
              ||||||||||||||||||||||||||||
Db        421 TSISSRKRGKRRYFWEYSEQLTPSQQER 448


RESULT 2
US-10-732-923-3319
; Sequence 3319, Application US/10732923
; Publication No. US20050108791A1
; GENERAL INFORMATION:
;  APPLICANT: Edgerton, Michael D
;  TITLE OF INVENTION: TRANSGENIC PLANTS WITH IMPROVED PHENOTYPES
;  FILE REFERENCE: 38-15(52796)C
;  CURRENT APPLICATION NUMBER: US/10/732,923
;  CURRENT FILING DATE:  2003-12-10
;  PRIOR APPLICATION NUMBER: 10/310,154
;  PRIOR FILING DATE: 2002-12-04
;  NUMBER OF SEQ ID NOS: 24149
; SEQ ID NO 3319
;   LENGTH: 1003
;   TYPE: PRT
;   ORGANISM: Kaposi's sarcoma-associated herpesvirus
US-10-732-923-3319

  Query Match             8.1%;  Score 191.5;  DB 5;  Length 1003;
  Best Local Similarity   24.7%;  Pred. No. 0.00028;
  Matches   92;  Conservative   53;  Mismatches  139;  Indels   89;  Gaps   14;

Qy        124 SSPYQDEEVISKTPTLAQLNSEDSQSVSDSLYYPDSLFSVKQNPLPSSFPGKKITSRAAA 183
              |||       :|   |        |: | | :|    |  | :: || ||  |:  :|
Db        111 SSPIPPSHPVSPGTT-------DTHSPSPALPPTQSPESSQRPPL-SSPTGRPDSSTPMR 162

Qy        184 PVCSSKTLQAEVPLSDCVQKASK------PPSSTQIMVKTNMYHNEKVNFHVECKDYVKK 237
              |  | :|     | :   :  ||       ||: : : |  :                 :
Db        163 PPPSQQTTPPHSPTTPPPEPPSKSSPDSLAPSTLRSLRKRRLSS--------------PQ 208

Qy        238 AKVKINPVQQSRPLL--------SQIHTDAAKENTCYCGAVAKRQEKKGMEPLQ----GH 285
                :||: || |:          ::   | |:  | |: :          |        |
Db        209 GPSTLNPICQSPPVSPPRCDFANRSVYPPWATESPIYVGSSSDGDTPPRQPPTSPISIGS 268

Qy        286 ATPALPF--KETQELLL--------SPLPQEGPGSLAAGESSSLSASTSVSDSSQKKEEH 335
              ::|:      :|  |:|        |   :|   :  |||    :    :|  ||  ::
Db        269 SSPSEGSWGDDTAMLVLLAEIAEEASKNEKECSENNQAGED---NGDNEISKESQVDKDD 325

Qy        336 NYSLFVSDNLGEQPTKCSPEEDEEDEE--------------DVDDEDHDEGFGSEHELS 380
              |      ||  ::  : : ||||||:|               : |||: ||    | :
Db        326 N------DNKDDEEEQETDEEDEEDDEEDDEEDDEEDDEEDDEEDDEEDDEEEDEE 379

Qy        381 ENEEEEEEEEDYEDDKDDDISDTFSEPGYENDSVEDLKE------VTSISSRKRGKRRYF 434
              |:||||:|||: ||::|||  |   |   | |  || ::        || | :: :
Db        380 EDEEEEDEEEEEEDEEDDDDEDNEDE---EEDKKEDEEDGGDGNKTLSIQSSQQQQE--- 433

Qy        435 WEYSEQLTPSQQE 447
                :|  | |||
Db        434 ---PQQQEPQQQE 443


RESULT 3
US-09-894-273-2
; Sequence 2, Application US/09894273
; Publication No. US20040037847A1
; GENERAL INFORMATION:
;  APPLICANT: Kieff, Elliott D.
;  APPLICANT:  Ballestas, Mary E.
;  APPLICANT:  Kaye, Kenneth M.
;  TITLE OF INVENTION: RHADINO VIRUS LANA ACTS IN TRANS ON A UNIT OF RHADINO
;  TITLE OF INVENTION:  VIRUS DNA TO MEDIATE EFFICIENT EPISOME PERSISTENCE
;  FILE REFERENCE: 16412-10001R
;  CURRENT APPLICATION NUMBER: US/09/894,273
;  CURRENT FILING DATE:  2001-06-28
;  PRIOR APPLICATION NUMBER: US 60/109,422
;  PRIOR FILING DATE: 1998-11-19
;  NUMBER OF SEQ ID NOS: 3
;  SOFTWARE: PatentIn Ver. 2.0
; SEQ ID NO 2
;   LENGTH: 1162
;   TYPE: PRT
;   ORGANISM: Kaposi's sarcoma-associated herpesvirus
US-09-894-273-2

  Query Match             8.0%;  Score 190;  DB 3;  Length 1162;
  Best Local Similarity   24.2%;  Pred. No. 0.00043;
  Matches   94;  Conservative   50;  Mismatches  137;  Indels  108;  Gaps   14;

Qy        124 SSPYQDEEVISKTPTLAQLNSEDSQSVSDSLYYPDSLFSVKQNPLPSSFPGKKITSRAAA 183
              |||       :|   |        |: | | :|    |  | :: || ||  |:  :|
Db        111 SSPIPPSHPVSPGTT-------DTHSPSPALPPTQSPESSQRPPL-SSPTGRPDSSTPMR 162

Qy        184 PVCSSKTLQAEVPLSDCVQKASK------PPSSTQIMVKTNMYHNEKVNFHVECKDYVKK 237
              |  | :|     | :   :  ||       ||: : : |  :                 :
Db        163 PPPSQQTTPPHSPTTPPPEPPSKSSPDSLAPSTLRSLRKRRLSS--------------PQ 208

Qy        238 AKVKINPVQQSRPLL--------SQIHTDAAKENTCYCGAVAKRQEKKGMEPLQ----GH 285
                :||: || |:          ::   | |:  | |: :          |        |
Db        209 GPSTLNPICQSPPVSPPRCDFANRSVYPPWATESPIYVGSSSDGDTPPRQPPTSPISIGS 268

Qy        286 ATPALPF--KETQELLL--------SPLPQEGPGSLAAGESSSLSASTSVSDSSQKKEEH 335
              ::|:      :|  |:|        |   :|   :  |||    :    :|  ||  ::
Db        269 SSPSEGSWGDDTAMLVLLAEIAEEASKNEKECSENNQAGED---NGDNEISKESQVDKDD 325

Qy        336 NYSLFVSDNLGEQPTKCSPEEDEEDEE--------------------------DVDDED 368
              |      ||  ::  : : ||||||:|                          : |||:
Db        326 N------DNKDDEEEQETDEEDEEDDEEDDEEDDEEDDEEDDEEDDEEDDEEDDEE 379

Qy        369 HDEGFGSEHELSENEEEEEEEEDYEDDKDDDISDTFSEPGYENDSVEDLKE--------- 419
              ||    | :  |:|||:||||| ||| |:|  |       |:|  || ||
Db        380 DDEEDDEEEDEEDEEEDEEEDEEDDDDEDNEDE------EDDEEEDKKEDEEDGGDGN 433

Qy        420 -VTSISSRKRGKRRYFWEYSEQLTPSQQE 447
               || | :: :        :|  | |||
Db        434 KTLSIQSSQQQQE------PQQQEPQQQE 456


RESULT 4
US-10-294-804-2
; Sequence 2, Application US/10294804
; Publication No. US20030133948A1
; GENERAL INFORMATION:
;  APPLICANT: Robertson, Erle S.
;  APPLICANT:  Cotter, Murray A.
;  TITLE OF INVENTION: Methods to Inhibit or Enhance the Binding of Viral DNA
;  TITLE OF INVENTION:  to Genomic Host DNA
;  FILE REFERENCE: UM-03778
;  CURRENT APPLICATION NUMBER: US/10/294,804
;  CURRENT FILING DATE:  2002-11-14
;  PRIOR APPLICATION NUMBER: US/09/410,399
;  PRIOR FILING DATE: 1999-10-01
;  NUMBER OF SEQ ID NOS: 6
;  SOFTWARE: PatentIn Ver. 2.0
; SEQ ID NO 2
;   LENGTH: 1162
;   TYPE: PRT
;   ORGANISM: Kaposi's sarcoma-associated herpesvirus
US-10-294-804-2

  Query Match             8.0%;  Score 190;  DB 4;  Length 1162;
  Best Local Similarity   24.2%;  Pred. No. 0.00043;
  Matches   94;  Conservative   50;  Mismatches  137;  Indels  108;  Gaps   14;

Qy        124 SSPYQDEEVISKTPTLAQLNSEDSQSVSDSLYYPDSLFSVKQNPLPSSFPGKKITSRAAA 183
              |||       :|   |        |: | | :|    |  | :: || ||  |:  :|
Db        111 SSPIPPSHPVSPGTT-------DTHSPSPALPPTQSPESSQRPPL-SSPTGRPDSSTPMR 162

Qy        184 PVCSSKTLQAEVPLSDCVQKASK------PPSSTQIMVKTNMYHNEKVNFHVECKDYVKK 237
```

```
Db        163 PPPSQQTTPPHSPTTPPPEPPSKSSPDSLAPSTLRSLRKRRLSS--------------PQ 208

Qy        238 AKVKINPVQQSRPLL--------SQIHTDAAKENTCYCGAVAKRQEKKGMEPLQ----GH 285

Db        209 GPSTLNPICQSPPVSPPRCDFANRSVYPPWATESPIYVGSSSDGDTPPRQPPTSPISIGS 268

Qy        286 ATPALPF--KETQELLL--------SPLPQEGPGSLAAGESSSLSASTSVSDSSQKKEEH 335

Db        269 SSPSEGSWGDDTAMLVLLAEIAEEASKNEKECSENNQAGED---NGDNEISKESQVDKDD 325

Qy        336 NYSLFVSDNLGEQPTKCSPEEDEEDEE----------------------------DVDDED 368

Db        326 N------DNKDDEEEQETDEEDEEDDEEDDEEDDEEDDEEDDEEDDEEDDEEDDEEDEE 379

Qy        369 HDEGFGSEHELSENEEEEEEEEDYEDDKDDDISDTFSEPGYENDSVEDLKE--------- 419

Db        380 DDEEDDEEEDEEEDEEEDEEEEDEEDDDDEDNEDE------EDDEEEDKKEDEEDGGDGN 433

Qy        420 -VTSISSRKRGKRRYFWEYSEQLTPSQQE 447

Db        434 KTLSIQSSQQQQE------PQQQEPQQQE 456


RESULT 5
US-10-194-046-2
; Sequence 2, Application US/10194046
; Publication No. US20040248081A1
; GENERAL INFORMATION:
;  APPLICANT: Robertson, Erle S.
;  TITLE OF INVENTION: LANA Viral Protein Binding Sites
;  FILE REFERENCE: UM-07142
;  CURRENT APPLICATION NUMBER: US/10/194,046
;  CURRENT FILING DATE:  2002-07-11
;  PRIOR APPLICATION NUMBER: 09/410,399
;  PRIOR FILING DATE: 1999-10-01
;  NUMBER OF SEQ ID NOS: 22
;  SOFTWARE: PatentIn version 3.1
; SEQ ID NO 2
;   LENGTH: 1162
;   TYPE: PRT
;   ORGANISM: Kaposi's sarcoma-associated herpesvirus
US-10-194-046-2

  Query Match             8.0%;  Score 190;  DB 5;  Length 1162;
  Best Local Similarity   24.2%;  Pred. No. 0.00043;
  Matches   94;  Conservative   50;  Mismatches  137;  Indels  108;  Gaps   14;

Qy        124 SSPYQDEEVISKTPTLAQLNSEDSQSVSDSLYYPDSLFSVKQNPLPSSFPGKKITSRAAA 183

Db        111 SSPIPPSHPVSPGTT-------DTHSPSPALPPTQSPESSQRPPL-SSPTGRPDSSTPMR 162

Qy        184 PVCSSKTLQAEVPLSDCVQKASK------PPSSTQIMVKTNMYHNEKVNFHVECKDYVKK 237

Db        163 PPPSQQTTPPHSPTTPPPEPPSKSSPDSLAPSTLRSLRKRRLSS--------------PQ 208

Qy        238 AKVKINPVQQSRPLL--------SQIHTDAAKENTCYCGAVAKRQEKKGMEPLQ----GH 285

Db        209 GPSTLNPICQSPPVSPPRCDFANRSVYPPWATESPIYVGSSSDGDTPPRQPPTSPISIGS 268

Qy        286 ATPALPF--KETQELLL--------SPLPQEGPGSLAAGESSSLSASTSVSDSSQKKEEH 335

Db        269 SSPSEGSWGDDTAMLVLLAEIAEEASKNEKECSENNQAGED---NGDNEISKESQVDKDD 325

Qy        336 NYSLFVSDNLGEQPTKCSPEEDEEDEE----------------------------DVDDED 368

Db        326 N------DNKDDEEEQETDEEDEEDDEEDDEEDDEEDDEEDDEEDDEEDDEEDDEEDEE 379

Qy        369 HDEGFGSEHELSENEEEEEEEEDYEDDKDDDISDTFSEPGYENDSVEDLKE--------- 419

Db        380 DDEEDDEEEDEEEDEEEDEEEEDEEDDDDEDNEDE------EDDEEEDKKEDEEDGGDGN 433

Qy        420 -VTSISSRKRGKRRYFWEYSEQLTPSQQE 447

Db        434 KTLSIQSSQQQQE------PQQQEPQQQE 456


RESULT 6
US-10-483-505-3
; Sequence 3, Application US/10483505
; Publication No. US20050118582A1
; GENERAL INFORMATION:
;  APPLICANT: INCYTE CORPORATION; SWARNAKAR, Anita;
;  APPLICANT:  RICHARDSON, Thomas W.; WARREN, Bridget A.;
;  APPLICANT:  GRIFFIN, Jennifer A.; TANG, Y. Tom;
;  APPLICANT:  YUE, Henry; BAUGHN, Mariah R.;
;  APPLICANT:  EMERLING, Brooke M.; LAL, Preeti G.;
;  APPLICANT:  LU, Dyung Aina M.; FORSYTHE, Ian J.;
;  APPLICANT:  RAMKUMAR, Jayalaxmi; LI, Joana X.;
;  APPLICANT:  BECHA, Shanya D.; DUGGAN, Brendan M.;
;  APPLICANT:  SANJANWALA, Madhusudan M.; LEE, Ernestine A.;
;  APPLICANT:  BURFORD, Neil; ELLIOTT, Vicki S.;
;  APPLICANT:  ISON, Craig H.; DING, Li;
;  APPLICANT:  BOROWSKY, Mark L.; YAO, Monique G.;
;  APPLICANT:  BARROSO, Ines; TRAN, Bao;
;  APPLICANT:  WALIA, Narinder K.; HAFALIA, April J. A.;
;  APPLICANT:  NGUYEN, Danniel B.; LU, Yan;
;  APPLICANT:  ARVIZU, Chandra S.
;  TITLE OF INVENTION: NUCLEIC ACID-ASSOCIATED PROTEINS
;  FILE REFERENCE: PF-1068 USN
;  CURRENT APPLICATION NUMBER: US/10/483,505
;  CURRENT FILING DATE:  2004-12-12
;  PRIOR APPLICATION NUMBER: PCT/US02/21971
;  PRIOR FILING DATE: 2002-07-10
;  PRIOR APPLICATION NUMBER: US 60/305,089
;  PRIOR FILING DATE: 2001-07-12
;  PRIOR APPLICATION NUMBER: US 60/305,104
;  PRIOR FILING DATE: 2001-07-12
;  PRIOR APPLICATION NUMBER: US 60/305,325
;  PRIOR FILING DATE: 2001-07-13
;  PRIOR APPLICATION NUMBER: US 60/305,390
;  PRIOR FILING DATE: 2001-07-13
;  PRIOR APPLICATION NUMBER: US 60/306,960
;  PRIOR FILING DATE: 2001-07-19
;  PRIOR APPLICATION NUMBER: US 60/306,694
;  PRIOR FILING DATE: 2001-07-20
;  PRIOR APPLICATION NUMBER: US 60/308,170
;  PRIOR FILING DATE: 2001-07-27
;  NUMBER OF SEQ ID NOS: 70
;  SOFTWARE: PERL Program
; SEQ ID NO 3
;   LENGTH: 735
;   TYPE: PRT
;   ORGANISM: Homo sapiens
;   FEATURE:
;   NAME/KEY: misc_feature
;   OTHER INFORMATION: Incyte ID No: 7037554CD1
US-10-483-505-3

  Query Match             7.2%;  Score 169.5;  DB 5;  Length 735;
  Best Local Similarity   24.5%;  Pred. No. 0.006;
  Matches   86;  Conservative   44;  Mismatches  120;  Indels  101;  Gaps   16;

Qy        120 LDDFSS--PYQDEEVISKTPTLAQLNSEDSQSVSDSLYYPDSLFSVKQNPLPSSFPGKKI 177

Db         16 LDDILTEVPEQDDELYNPESEQDKNEKKGSKRKSDRM---ESTDTKRQKP---SVHSRQL 69

Qy        178 TSRAAAPVCSSKTLQAEVPLSDCVQK-----ASKPPSSTQIMVKTNMYHNEKVNFHVECK 232

Db         70 VSK--------------PLSSSVSNNKRIVSTKGKSATE-------YKNEEYQRSERNK 107

Qy        233 DYVKKAKVKINPVQQSRPLLSQIHTDAAKENTCYCGAVAKRQEKKGMEPLQGHATPALPF 292

Db        108 RLDADRKIRLSSSASREPYKNQ------PEKTC----VRKRDPERRAK------------ 145

Qy        293 KETQELLLSPLPQEGPGSLAAGESSSLSASTSVSDSSQKKEEHNY-------SLFVSDNL 345
```

```
                    || |    ||    |       || |    | ::      |        |    ||
Db        146 --------SPTPD---GSERIGLEVDRRASRSSQSSKEEVNSEEYGSDHETGSSGSSDEQ 194

Qy        346 G------EQPTKCSPEEDEEDEEDVD-----DEDHDEGFGSEHELSENEEEEEEEEDYED 394
              |       |:  :   ||||| ||| :     ||| :|    | |   | |||||||:||
Db        195 GNNTENEEEGVEEDVEEDEEVEEDAEEDEEVDEDGEE--EEEEEEEEEEEEEEEEEYEQ 252

Qy        395 DKDD------------DISDTFSEP-GYENDSVEDLKEVTSISSRKRGKRR 432
              |: |            : ||: ||   : : ||           :|: ::|
Db        253 DERDQKEEGNDYDTRSEASDSGSESVSFTDGSVRSGSGTDGSDEKKKERKR 303


RESULT 7
US-09-825-886-22
; Sequence 22, Application US/09825886
; Publication No. US20020076693A1
; GENERAL INFORMATION:
;  APPLICANT: Hovanessian, Ara
;  APPLICANT:  Callebaut, Christian
;  APPLICANT:  Krust, Bernard
;  APPLICANT:  Jacotot, Etienne
;  APPLICANT:  Muller, Sylviane
;  APPLICANT:  Briand, Jean-Paul
;  APPLICANT:  Guichard, Giles
;  TITLE OF INVENTION: A NOVEL CELL SURFACE RECEPTOR FOR HIV RETROVIRUSES,
;  TITLE OF INVENTION:  THERAPEUTIC AND DIAGNOSTIC USES.
;  FILE REFERENCE: 03495.0166-01000
;  CURRENT APPLICATION NUMBER: US/09/825,886
;  CURRENT FILING DATE:  2001-07-26
;  PRIOR APPLICATION NUMBER: 09/393,302
;  PRIOR FILING DATE: 1999-09-10
;  PRIOR APPLICATION NUMBER: PCT/EP98/01409
;  PRIOR FILING DATE: 1998-03-12
;  PRIOR APPLICATION NUMBER: 60/040,969
;  PRIOR FILING DATE: 1997-03-12
;  NUMBER OF SEQ ID NOS: 32
;  SOFTWARE: PatentIn Ver. 2.1
; SEQ ID NO 22
;   LENGTH: 707
;   TYPE: PRT
;   ORGANISM: Homo sapiens
US-09-825-886-22

  Query Match             7.0%;  Score 165;  DB 3;  Length 707;
  Best Local Similarity   22.0%;  Pred. No. 0.012;
  Matches   76;  Conservative   52;  Mismatches  123;  Indels   94;  Gaps   12;

Qy        129 DEEVISKTPTLAQLNSEDSQSVSDSLYYPDSLFSVKQNPLPSSFPGKKITSRAAAPVCSS 188
              | : ::  |   : :||| :   |      ||  | ::  :|    |||  : :|  |  |
Db         13 DPKKMAPPPKEVEEDSEDEEMSEDE--EDDS--SGEEVVIPQK-KGKKAAATSAKKVVVS 67

Qy        189 KTLQAEVPLSDCVQKASKPPSSTQIMVKTNMYHNEKVNFHVECKDYVKKAKVKINPVQQS 248
              |   :| ::   :||:  |                                |      || : | :
Db         68 PT--KKVAVATPAKKAAVTPGK--------------------KAAATPAKKTVTPAK-- 102

Qy        249 RPLLSQIHTDAAKENTCYCGAVAKRQEKKGMEPLQGHATPALPFKETQELLLSPLPQEGP 308
                                        |||  |  | |  | | |:      : :| :|
Db        103 ---------------------AVTTPGKKGATP--GKALVATPGKKG-----AAIPAKGA 134

Qy        309 GSLAAGESSSLSASTSVSDSSQKKEEHNYSLFVSDNLGEQP--------TKCSPEEDEED 360
               :   |:::    |    |   :::| :      |    :|           | :||:||
Db        135 KN---GKNAKKEDSDEEEDDDSEEDEEDDEDEDEDEDEIEPAAMKAAAAAPASEDEDDED 191

Qy        361 EEDVDDEDHDEGFGSEHE---------------------LSENEEEEEEEEDYEDDKDD 398
              :|| :|:| ||   || |                   ::|:|:|:::|:|| :|| |:
Db        192 DEDDEDDDDDEEDDSEEEAMETTPAKGKKAAKVVPVKAKNVAEDEDEEEDDEDEDDDDDE 251

Qy        399 DISDTFSEPGYENDSVED---LKEVTSISSRKRGKRRYFWEYSEQ 440
              |  |   |   | :  |:   :||      ::  |::   |  :|
Db        252 DDEDDDDEDDEEEEEEEEEPVKEAPGKRKKEMAKQKAAPEAKKQ 296


RESULT 8
US-10-384-569-3
; Sequence 3, Application US/10384569
; Publication No. US20040002457A1
; GENERAL INFORMATION:
;  APPLICANT: HOVANESSIAN, ARA G.
;  APPLICANT:  BRIAND, JEAN-PAUL
;  TITLE OF INVENTION: INHIBITION OF VIRUS ANCHORAGE BY RGG DOMAIN OF A CELL SURFACE-
;  TITLE OF INVENTION:  EXPRESSED PROTEIN, POLYNUCLEOTIDE CODING FOR SAID RGG DOMAIN,
;  TITLE OF INVENTION:  THERAPEUTIC USES THEREOF BY INHIBITION OF MICROORGANISM OR
;  TITLE OF INVENTION:  PROTEIN LIGAND BINDING TO THE CELL-SURFACE-EXPRESSED PROTEIN
;  FILE REFERENCE: 03495.0260-00000
;  CURRENT APPLICATION NUMBER: US/10/384,569
;  CURRENT FILING DATE:  2004-07-14
;  PRIOR APPLICATION NUMBER: 60/363,371
;  PRIOR FILING DATE: 2002-03-12
;  PRIOR APPLICATION NUMBER: 60/397,600
;  PRIOR FILING DATE: 2002-07-23
;  NUMBER OF SEQ ID NOS: 25
;  SOFTWARE: PatentIn Ver. 2.1
; SEQ ID NO 3
;   LENGTH: 707
;   TYPE: PRT
;   ORGANISM: Homo sapiens
US-10-384-569-3

  Query Match             7.0%;  Score 165;  DB 4;  Length 707;
  Best Local Similarity   22.0%;  Pred. No. 0.012;
  Matches   76;  Conservative   52;  Mismatches  123;  Indels   94;  Gaps   12;

Qy        129 DEEVISKTPTLAQLNSEDSQSVSDSLYYPDSLFSVKQNPLPSSFPGKKITSRAAAPVCSS 188
              | : ::  |   : :||| :   |      ||  | ::  :|    |||  : :|  |  |
Db         13 DPKKMAPPPKEVEEDSEDEEMSEDE--EDDS--SGEEVVIPQK-KGKKAAATSAKKVVVS 67

Qy        189 KTLQAEVPLSDCVQKASKPPSSTQIMVKTNMYHNEKVNFHVECKDYVKKAKVKINPVQQS 248
              |   :| ::   :||:  |                                |      || : | :
Db         68 PT--KKVAVATPAKKAAVTPGK--------------------KAAATPAKKTVTPAK-- 102

Qy        249 RPLLSQIHTDAAKENTCYCGAVAKRQEKKGMEPLQGHATPALPFKETQELLLSPLPQEGP 308
                                        |||  |  | |  | | |:      : :| :|
Db        103 ---------------------AVTTPGKKGATP--GKALVATPGKKG-----AAIPAKGA 134

Qy        309 GSLAAGESSSLSASTSVSDSSQKKEEHNYSLFVSDNLGEQP--------TKCSPEEDEED 360
               :   |:::    |    |   :::| :      |    :|           | :||:||
Db        135 KN---GKNAKKEDSDEEEDDDSEEDEEDDEDEDEDEDEIEPAAMKAAAAAPASEDEDDED 191

Qy        361 EEDVDDEDHDEGFGSEHE---------------------LSENEEEEEEEEDYEDDKDD 398
              :|| :|:| ||   || |                   ::|:|:|:::|:|| :|| |:
Db        192 DEDDEDDDDDEEDDSEEEAMETTPAKGKKAAKVVPVKAKNVAEDEDEEEDDEDEDDDDDE 251

Qy        399 DISDTFSEPGYENDSVED---LKEVTSISSRKRGKRRYFWEYSEQ 440
              |  |   |   | :  |:   :||      ::  |::   |  :|
Db        252 DDEDDDDEDDEEEEEEEEEPVKEAPGKRKKEMAKQKAAPEAKKQ 296


RESULT 9
US-10-400-083-19
; Sequence 19, Application US/10400083
; Publication No. US20040186056A1
; GENERAL INFORMATION:
;  APPLICANT: Ruoslahti, Erkki
;  APPLICANT:  Porkka, Kimmo
;  APPLICANT:  Christian, Sven
;  TITLE OF INVENTION: HMGN2 Peptides and Related Molecules
;  TITLE OF INVENTION:  that Selectively Home to Tumor Blood Vessels and Tumor Cells
;  FILE REFERENCE: P-LJ 5662
;  CURRENT APPLICATION NUMBER: US/10/400,083
;  CURRENT FILING DATE:  2003-03-20
;  PRIOR APPLICATION NUMBER: US 10/116,866
;  PRIOR FILING DATE: 2002-04-05
;  NUMBER OF SEQ ID NOS: 20
;  SOFTWARE: FastSEQ for Windows Version 4.0
; SEQ ID NO 19
;   LENGTH: 707
```

```
;   TYPE: PRT
;   ORGANISM: Homo sapiens
US-10-400-083-19

  Query Match             7.0%;  Score 165;  DB 4;  Length 707;
  Best Local Similarity   22.0%;  Pred. No. 0.012;
  Matches   76;  Conservative   52;  Mismatches  123;  Indels   94;  Gaps   12;

Qy        129 DEEVISKTPTLAQLNSEDSQSVSDSLYYPDSLFSVKQNPLPSSFPGKKITSRAAAPVCSS 188
              | : ::  |   : :||| :    |     ||  | ::  :|    |||  : :|  |  |
Db         13 DPKKMAPPPKEVEEDSEDEEMSEDE--EDDS--SGEEVVIPQK-KGKKAAATSAKKVVVS 67

Qy        189 KTLQAEVPLSDCVQKASKPPSSTQIMVKTNMYHNEKVNFHVECKDYVKKAKVKINPVQQS 248
                |   :| ::   :||:  |                        |     ||  : | :
Db         68 PT--KKVAVATPAKKAAVTPGK--------------------KAAATPAKKTVTPAK-- 102

Qy        249 RPLLSQIHTDAAKENTCYCGAVAKRQEKKGMEPLQGHATPALPFKETQELLLSPLPQEGP 308
                                    |||  |  | |  | | |:      : :| :|
Db        103 --------------------AVTTPGKKGATP--GKALVATPGKKG-----AAIPAKGA 134

Qy        309 GSLAAGESSSLSASTSVSDSSQKKEEHNYSLFVSDNLGEQP--------TKCSPEEDEED 360
                :   |:::    |    |   :::| :      |    :|           | :||:||
Db        135 KN---GKNAKKEDSDEEEDDDSEEDEEDDEDEDEDEDEIEPAAMKAAAAAPASEDEDDED 191

Qy        361 EEDVDDEDHDEGFGSEHE---------------------LSENEEEEEEEEDYEDDKDD 398
              :|| :|:| ||   || |                    ::|:|:|||::|| :|| |:
Db        192 DEDDEDDDDDEEDDSEEEAMETTPAKGKKAAKVVPVKAKNVAEDEDEEEDDEDEDDDDDE 251

Qy        399 DISDTFSEPGYENDSVED---LKEVTSISSRKRGKRRYFWEYSEQ 440
              | |   |  | :  |:   :||      ::  |::   |  :|
Db        252 DDEDDDDEDDEEEEEEEEEPVKEAPGKRKKEMAKQKAAPEAKKQ 296


RESULT 10
US-10-491-545A-20
; Sequence 20, Application US/10491545A
; Publication No. US20050130117A1
; GENERAL INFORMATION:
;  APPLICANT: Chu, Peter
;  APPLICANT:  Li, Congfen
;  APPLICANT:  Liao, X. Charlene
;  APPLICANT:  Masuda, Esteban
;  APPLICANT:  Pardo, Jorge
;  APPLICANT:  Zhao, Haoran
;  APPLICANT:  Rigel Pharmaceuticals, Inc.
;  TITLE OF INVENTION: Modulators of Lymphocyte Activation and Migration
;  FILE REFERENCE: 021044-000330US
;  CURRENT APPLICATION NUMBER: US/10/491,545A
;  CURRENT FILING DATE:  2004-04-01
;  PRIOR APPLICATION NUMBER: US 60/327,212
;  PRIOR FILING DATE: 2001-10-03
;  PRIOR APPLICATION NUMBER: WO PCT/US02/31618
;  PRIOR FILING DATE: 2002-10-02
;  NUMBER OF SEQ ID NOS: 96
;  SOFTWARE: PatentIn Ver. 2.1
; SEQ ID NO 20
;   LENGTH: 707
;   TYPE: PRT
;   ORGANISM: Homo sapiens
;   FEATURE:
;   OTHER INFORMATION: human nucleolin (NCL)
US-10-491-545A-20

  Query Match             7.0%;  Score 165;  DB 5;  Length 707;
  Best Local Similarity   22.0%;  Pred. No. 0.012;
  Matches   76;  Conservative   52;  Mismatches  123;  Indels   94;  Gaps   12;

Qy        129 DEEVISKTPTLAQLNSEDSQSVSDSLYYPDSLFSVKQNPLPSSFPGKKITSRAAAPVCSS 188
              | : ::  |   : :||| :    |     ||  | ::  :|    |||  : :|  |  |
Db         13 DPKKMAPPPKEVEEDSEDEEMSEDE--EDDS--SGEEVVIPQK-KGKKAAATSAKKVVVS 67

Qy        189 KTLQAEVPLSDCVQKASKPPSSTQIMVKTNMYHNEKVNFHVECKDYVKKAKVKINPVQQS 248
                |   :| ::   :||:  |                        |     ||  : | :
Db         68 PT--KKVAVATPAKKAAVTPGK--------------------KAAATPAKKTVTPAK-- 102

Qy        249 RPLLSQIHTDAAKENTCYCGAVAKRQEKKGMEPLQGHATPALPFKETQELLLSPLPQEGP 308
                                    |||  |  | |  | | |:      : :| :|
Db        103 --------------------AVTTPGKKGATP--GKALVATPGKKG-----AAIPAKGA 134

Qy        309 GSLAAGESSSLSASTSVSDSSQKKEEHNYSLFVSDNLGEQP--------TKCSPEEDEED 360
                :   |:::    |    |   :::| :      |    :|           | :||:||
Db        135 KN---GKNAKKEDSDEEEDDDSEEDEEDDEDEDEDEDEIEPAAMKAAAAAPASEDEDDED 191

Qy        361 EEDVDDEDHDEGFGSEHE---------------------LSENEEEEEEEEDYEDDKDD 398
              :|| :|:| ||   || |                    ::|:|:|||::|| :|| |:
Db        192 DEDDEDDDDDEEDDSEEEAMETTPAKGKKAAKVVPVKAKNVAEDEDEEEDDEDEDDDDDE 251

Qy        399 DISDTFSEPGYENDSVED---LKEVTSISSRKRGKRRYFWEYSEQ 440
              | |   |  | :  |:   :||      ::  |::   |  :|
Db        252 DDEDDDDEDDEEEEEEEEEPVKEAPGKRKKEMAKQKAAPEAKKQ 296


RESULT 11
US-10-450-763-41629
; Sequence 41629, Application US/10450763
; Publication No. US20050196754A1
; GENERAL INFORMATION:
;  APPLICANT: Hyseq, Inc
;  TITLE OF INVENTION: NOVEL NUCLEIC ACIDS AND POLYPEPTIDES
;  FILE REFERENCE: 790CIP3/US
;  CURRENT APPLICATION NUMBER: US/10/450,763
;  CURRENT FILING DATE:  2003-06-11
;  PRIOR APPLICATION NUMBER: PCT/US01/08631
;  PRIOR FILING DATE: 2001-03-30
;  PRIOR APPLICATION NUMBER: 09/540,217
;  PRIOR FILING DATE: 2000-03-31
;  PRIOR APPLICATION NUMBER: 09/649,167
;  PRIOR FILING DATE: 2000-08-23
;  NUMBER OF SEQ ID NOS: 60736
;  SOFTWARE: Custom
; SEQ ID NO 41629
;   LENGTH: 167
;   TYPE: PRT
;   ORGANISM: Homo sapiens
;   FEATURE:
;   NAME/KEY: DOMAIN
;   LOCATION: (117)..(167)
;   OTHER INFORMATION: Neuromodulin (GAP-43) proteins domain identified by eMATRIX,
;   OTHER INFORMATION: accession number BL00412D, p-value=8.644e-15, raw score of 16.54
US-10-450-763-41629

  Query Match             6.9%;  Score 162;  DB 5;  Length 167;
  Best Local Similarity   30.5%;  Pred. No. 0.0029;
  Matches   46;  Conservative   28;  Mismatches   55;  Indels   22;  Gaps    3;

Qy        281 PLQGHATPALPFKETQELLLSPLPQ-----------EGPGSLAAGESSSLSASTSVSDS 328
              |   || |  |   : ||    |||             ||  | :   :||||  :  |
Db          3 PGDRHAFPLAPSGLSPEL---TLPQTQCCAQATVQGLEGTRSWSQSGTSSLSPWSHTSLR 59

Qy        329 SQKKEEHNYSLFVSDNLGEQPTKCSPEEDEEDEEDVDDEDHDEGFGSEHELSENEEEEEE 388
                ::|||        :   |:  :   ||:||:||: ::|: :|    | |: | ||||||
Db         60 RRRKEE-------GEEEEEEEEEEEEEEEEEEEEEEEEEEEEEEEEEEEEIEEGEEEEEE 112

Qy        389 EEDYEDDKDDDISDTFSEPGYENDSVEDLKE 419
              ||: |:::::  :   |   | :  |: :|
Db        113 EEEEEEEEEEEEEEEEEEEEEEEEEEEEE 143


RESULT 12
US-10-450-763-34710
; Sequence 34710, Application US/10450763
; Publication No. US20050196754A1
; GENERAL INFORMATION:
;  APPLICANT: Hyseq, Inc
;  TITLE OF INVENTION: NOVEL NUCLEIC ACIDS AND POLYPEPTIDES
;  FILE REFERENCE: 790CIP3/US
```

```
;  CURRENT APPLICATION NUMBER: US/10/450,763
;  CURRENT FILING DATE:  2003-06-11
;  PRIOR APPLICATION NUMBER: PCT/US01/08631
;  PRIOR FILING DATE: 2001-03-30
;  PRIOR APPLICATION NUMBER: 09/540,217
;  PRIOR FILING DATE: 2000-03-31
;  PRIOR APPLICATION NUMBER: 09/649,167
;  PRIOR FILING DATE: 2000-08-23
;  NUMBER OF SEQ ID NOS: 60736
;  SOFTWARE: Custom
; SEQ ID NO 34710
;   LENGTH: 382
;   TYPE: PRT
;   ORGANISM: Homo sapiens
;   FEATURE:
;   NAME/KEY: DOMAIN
;   LOCATION: (308)..(343)
;   OTHER INFORMATION: Matrix protein (MA) p15. domain identified by eMATRIX,
;   OTHER INFORMATION: accession number PF01140D, p-value=2.570e-09, raw score of 15.54
US-10-450-763-34710

  Query Match             6.8%;  Score 160.5;  DB 5;  Length 382;
  Best Local Similarity   20.2%;  Pred. No. 0.011;
  Matches   86;  Conservative   58;  Mismatches  125;  Indels  157;  Gaps   16;

Qy         29 STDLLANSSDPDFMYELDREMNYQQNPRDNFLSLEDCKDIENL--ESFTDVLDNEGALTS 86
              | : ||| :    : : |  : ::   |    |   :||:    |   ||::    |
Db         99 SVEKLANGNIA--LVDSLRSRSLEEGDSDPHKRLSGAQDIKTTVEEVIADVVEIARELEL 156

Qy         87 NWEQWDTYCEDLTKYTK-----LTSCDIWGTKEVDYLGLDDFSSPYQDEEVISKTPTLAQ 141
                 |      ||:|:: :     ||  :::   |      |:  |:| :|  :|::|
Db        157 EVEP-----EDVTEFLQAHEKTLTDVELFLINEQIKWFLEMKSTPREDAVIIAET----- 206

Qy        142 LNSEDSQSVSDSLYYPDSLFSVKQNPLPSSFPGKKITSRAAAPVCSSKTLQAEVPLSDCV 201
                    ::  | |  :| : :|  :           |  |:      |  :    |
Db        207 --------ITKVLEYDINLVTKQQQGM----------RQLTPILKEVLLWVK-----CH 242

Qy        202 QKASKPPSSTQIMVKTNMYHNEKVNFHVECKDYVKKAKVKINPVQQSRPLLSQIHTDAAK 261
              | |                          |   | ::|    :|||         :|||
Db        243 QTA---------------------LHATEKPFIKG---RINP-------CGKIHTCLNL 270

Qy        262 ENTCYCGAVAKRQEKKGMEPLQGHATPALPFKETQELLLSPLPQEGPGSLAAGESSSLSA 321
              |   || :  |:|
Db        271 RN---CGQLLIREE---------------------------------------------- 281

Qy        322 STSVSDSSQKKEEHNYSLFVSDNLGEQPTKCSPEEDEEDEEDVDDEDHDEGFGSEHELSE 381
                  :   |:||  |                ||:||:||: ::|: :|    | |  |
Db        282 -----EEEDKEEEEQYE-------------EKEEEEEEEEEEEEEEEE----EEEEEE 318

Qy        382 NEEEEEEEEDYEDDKDDDISDTFSEPGYENDSVEDLKEVTSISSRKRGKRRYFWEYSEQL 441
              |||||||:: |:||:|:  |       | :  || ||       :  | :   |  |:
Db        319 EEEEEEEEDEEEEDKEDEEEDK------EKEEEEDNKE----EEEEEDKEKEEEEDKEEK 368

Qy        442 TPSQQE 447
                ::|
Db        369 EDKEEE 374


RESULT 13
US-10-369-493-5455
; Sequence 5455, Application US/10369493
; Publication No. US20030233675A1
; GENERAL INFORMATION:
;  APPLICANT: Cao, Yongwei
;  APPLICANT:  Hinkle, Gregory J.
;  APPLICANT:  Slater, Steven C.
;  APPLICANT:  Goldman, Barry S.
;  APPLICANT:  Chen, Xianfeng
;  TITLE OF INVENTION: EXPRESSION OF MICROBIAL PROTEINS IN PLANTS FOR PRODUCTION OF
;  TITLE OF INVENTION:  PLANTS WITH IMPROVED PROPERTIES
;  FILE REFERENCE: 38-10(52052)B
;  CURRENT APPLICATION NUMBER: US/10/369,493
;  CURRENT FILING DATE:  2003-02-28
;  PRIOR APPLICATION NUMBER: US 60/360,039
;  PRIOR FILING DATE: 2002-02-21
;  NUMBER OF SEQ ID NOS: 47374
; SEQ ID NO 5455
;   LENGTH: 1974
;   TYPE: PRT
;   ORGANISM: Caenorhabditis elegans
US-10-369-493-5455

  Query Match             6.8%;  Score 160.5;  DB 4;  Length 1974;
  Best Local Similarity   21.6%;  Pred. No. 0.088;
  Matches  104;  Conservative   51;  Mismatches  152;  Indels  175;  Gaps   19;

Qy         54 NPRDNFLSLEDCKDIENLE-------SFTDVLDNEGALTSNWEQWDTYCEDLTKYTKLTS 106
              | : | |  |: |  ||:        | |||    |
Db       1400 NDQINQLKQENLKLTENIRMAQLQSVSATDVESKPGP---------------------- 1436

Qy        107 CDIWGTKEVDYLGLDDFSSPYQDEEVISKTPT--LAQLNSEDSQSVSDSLYYPDSLFSVK 164
                                ||  :    | :|||  |  |:|    |         ||    :|
Db       1437 -----------------SSASKSVSSIRQTPTKVLDPLSSAAKQPNE-----PDQTTGLK 1474

Qy        165 QNPLPSSFPGKKITSRAAAPVCSSKTLQAEVPLSDCVQKASK-----PPS---------- 209
                :  | ||  |: :     |   |:    |  |     || |       |||
Db       1475 TSQQPPSFAAKRPSFGRQAQSSSNVIPPAPAPTPTSQQKVSPVKRPIPPSIPNEPLDIIP 1534

Qy        210 ---STQI---MVKTNMY-------HNEKVNFHVECKD-YVKKAKVKINPVQQSRPLLSQI 255
                  |   |    || :       |  : :  |  : :  :   : |  : : :| |
Db       1535 PVPSDNIPDPTPPTNSFGTVLPVPHTFQTSVRVPTQSLFSSSSTTTVQPQPEKKNVLPSI 1594

Qy        256 HTDAAKENTCYCGAVAKRQEKKGMEPLQ------GHATPALPFKETQELLLSPLPQEGPG 309
                |:|       |  :       | | |      |:    :|   |       ||:|
Db       1595 --DSAPSTP--GGNSSMVTTTSSMAPGQSIFGNIGN----VPVPTTAPTDNLALPEESVI 1646

Qy        310 SLAAGESSSLSAS-------------------------TSVSDSSQKKEEHNYSLFVSD 343
                :||:|| :| |                         | ||  :|:  ::: |  :
Db       1647 EGSAGQSSLVSGSIDQRKVQDIDLVANDGESRDSTNVGGVSSSDVRRKRTANDFELSEAK 1706

Qy        344 NLGEQPTKC------SPEEDEED-----------------------------------E 361
                | | | :        || |::|                                   |
Db       1707 RLRESPNETVTSIADIPELDDDDGVLGMEHEVSDEDPNDNTIQEQRPDVIDLENDEEVLE 1766

Qy        362 EDVDDEDHDEGFGSEHELSENEEEEEEEEDYEDDKDDDI---SDTFSEPGYENDSVEDLK 418
              :::|:|: |: ||::      || |:|||  ||| |||:   ||   ||  :||  | |
Db       1767 DEMDEEEDDDSFGND------EEFEDEEEIPEDDDDDVVVLSDGDDEPANDNDE-ESLN 1819

Qy        419 EV 420
              ::
Db       1820 DI 1821


RESULT 14
US-09-839-479-21
; Sequence 21, Application US/09839479
; Publication No. US20020039779A1
; GENERAL INFORMATION:
;  APPLICANT: Jones, Michael H.
;  TITLE OF INVENTION: TRANSCRIPTIONAL REGULATOR
;  FILE REFERENCE: 06501-042002
;  CURRENT APPLICATION NUMBER: US/09/839,479
;  CURRENT FILING DATE:  2001-04-20
;  PRIOR APPLICATION NUMBER: US 09/418,710
;  PRIOR FILING DATE: 1999-10-15
;  PRIOR APPLICATION NUMBER: PCT/JP98/01783
;  PRIOR FILING DATE: 1998-04-17
;  PRIOR APPLICATION NUMBER: JP 9/310027
;  PRIOR FILING DATE: 1997-10-24
;  PRIOR APPLICATION NUMBER: JP 9/116570
;  PRIOR FILING DATE: 1997-04-18
;  NUMBER OF SEQ ID NOS: 72
;  SOFTWARE: FastSEQ for Windows Version 4.0
; SEQ ID NO 21
;   LENGTH: 1972
;   TYPE: PRT
```

```
;   ORGANISM: Homo sapiens
US-09-839-479-21

  Query Match             6.8%;  Score 160;  DB 3;  Length 1972;
  Best Local Similarity   21.5%;  Pred. No. 0.095;
  Matches   94;  Conservative   68;  Mismatches  130;  Indels  146;  Gaps   20;

Qy        119 GLDDFSSPYQDEEVISKTPTLAQLNSEDSQSVSDSLYYPDSLFSVKQNPLPSSFPGKKIT 178
              |:    |   :|:   :  :: :   :|| | |:: :  ::   :      | :  |:| |
Db         63 GISSSDSDDLEEDEEEEDQSIEESEDDDSDSESEAQHKSNNQVLLHGISDPKA-DGQKAT 121

Qy        179 SRAAAPVCSSKTLQAEVPL-----SDCVQKASKPPS--STQI--MVKTNMYHNEKVNFH- 228
               :|       | :   :||     :   |   | |   | |:  : :::    | || |
Db        122 EKA-----QEKRIHQPLPLAFESQTHSFQSQQKQPQVLSQQLPFIFQSSQAKEESVNKHT 176

Qy        229 -----------VECKDYVKKAK----VKI------------NPVQQSRPLLSQIHTDAA 260
                        |:    | :||    :|:              |  ::|  | |:: :
Db        177 SVIQSTGLVSNVKPLSLVNQAKKETYMKLIVPSPDVLKAGNKNTSEESSLLTSELRSKRE 236

Qy        261 KENTCYCGAVAKRQEKKGME--------PLQGHATPALPFKETQE-----------LLLS 301
              :    :  : |::  | ::        |   ::|| | |:| |            || :
Db        237 QYKQAFPSQLKKQESSKSLKKVIAALSNPKATSSSPAHP-KQTLENNHPNPFLTNALLGN 295

Qy        302 PLP--------QEGP-------------GSLAAGESSSLSASTSVSDS----------- 328
              |         || |                ::||  |:  |:  ::| |
Db        296 HQPNGVIQSVIQEAPLALTTKTKMQSKINENIAAASSTPFSSPVNLSTSGRRTPGNQTPV 355

Qy        329 ---------SQKKEEHNYSLFVSDNLG-------EQPTKC-------------------- 352
                      || ||:      ||:|:          | |
Db        356 MPSASPILHSQGKEK-----AVSNNVNPVKTQHHSHPAKSLVEQFRGTDSDIPSSKDSED 410

Qy        353 ---SPEEDEEDEEDVDDEDHDEGFGSEHELSENEEEEEE--EEDYEDDKDDDISDTFSEP 407
                  |||:|:|:: |||| ||   |: |   | | : |  ||: :|||| | ||: :|
Db        411 SNEDEEEDDEEEDEEDDED-DESDDSQSESDSNSESDTEGSEEEDDDDKDQDESDSDTEG 469

Qy        408 GYENDSVEDLKEVTSISS 425
                |  |::  |  :|: |
Db        470 --EKTSMKLNKTTSSVKS 485


RESULT 15
US-10-376-537-21
; Sequence 21, Application US/10376537
; Publication No. US20030224405A1
; GENERAL INFORMATION:
;  APPLICANT: Jones, Michael H.
;  TITLE OF INVENTION: TRANSCRIPTIONAL REGULATOR
;  FILE REFERENCE: 06501-042001
;  CURRENT APPLICATION NUMBER: US/10/376,537
;  CURRENT FILING DATE:  2003-02-28
;  PRIOR APPLICATION NUMBER: US/09/418,710
;  PRIOR FILING DATE: 1999-10-15
;  PRIOR APPLICATION NUMBER: PCT/JP98/01783
;  PRIOR FILING DATE: 1998-04-17
;  PRIOR APPLICATION NUMBER: JP 9/310027
;  PRIOR FILING DATE: 1997-10-24
;  PRIOR APPLICATION NUMBER: JP 9/116570
;  PRIOR FILING DATE: 1997-04-18
;  NUMBER OF SEQ ID NOS: 73
;  SOFTWARE: FastSEQ for Windows Version 4.0
; SEQ ID NO 21
;   LENGTH: 1972
;   TYPE: PRT
;   ORGANISM: Homo sapiens
US-10-376-537-21

  Query Match             6.8%;  Score 160;  DB 4;  Length 1972;
  Best Local Similarity   21.5%;  Pred. No. 0.095;
  Matches   94;  Conservative   68;  Mismatches  130;  Indels  146;  Gaps   20;

Qy        119 GLDDFSSPYQDEEVISKTPTLAQLNSEDSQSVSDSLYYPDSLFSVKQNPLPSSFPGKKIT 178
              |:    |   :|:   :  :: :   :|| | |:: :  ::   :      | :  |:| |
Db         63 GISSSDSDDLEEDEEEEDQSIEESEDDDSDSESEAQHKSNNQVLLHGISDPKA-DGQKAT 121

Qy        179 SRAAAPVCSSKTLQAEVPL-----SDCVQKASKPPS--STQI--MVKTNMYHNEKVNFH- 228
               :|       | :   :||     :   |   | |   | |:  : :::    | || |
Db        122 EKA-----QEKRIHQPLPLAFESQTHSFQSQQKQPQVLSQQLPFIFQSSQAKEESVNKHT 176

Qy        229 -----------VECKDYVKKAK----VKI------------NPVQQSRPLLSQIHTDAA 260
                        |:    | :||    :|:              |  ::|  | |:: :
Db        177 SVIQSTGLVSNVKPLSLVNQAKKETYMKLIVPSPDVLKAGNKNTSEESSLLTSELRSKRE 236

Qy        261 KENTCYCGAVAKRQEKKGME--------PLQGHATPALPFKETQE-----------LLLS 301
              :    :  : |::  | ::        |   ::|| | |:| |            || :
Db        237 QYKQAFPSQLKKQESSKSLKKVIAALSNPKATSSSPAHP-KQTLENNHPNPFLTNALLGN 295

Qy        302 PLP--------QEGP-------------GSLAAGESSSLSASTSVSDS----------- 328
              |         || |                ::||  |:  |:  ::| |
Db        296 HQPNGVIQSVIQEAPLALTTKTKMQSKINENIAAASSTPFSSPVNLSTSGRRTPGNQTPV 355

Qy        329 ---------SQKKEEHNYSLFVSDNLG-------EQPTKC-------------------- 352
                      || ||:      ||:|:          | |
Db        356 MPSASPILHSQGKEK-----AVSNNVNPVKTQHHSHPAKSLVEQFRGTDSDIPSSKDSED 410

Qy        353 ---SPEEDEEDEEDVDDEDHDEGFGSEHELSENEEEEEE--EEDYEDDKDDDISDTFSEP 407
                  |||:|:|:: |||| ||   |: |   | | : |  ||: :|||| | ||: :|
Db        411 SNEDEEEDDEEEDEEDDED-DESDDSQSESDSNSESDTEGSEEEDDDDKDQDESDSDTEG 469

Qy        408 GYENDSVEDLKEVTSISS 425
                |  |::  |  :|: |
Db        470 --EKTSMKLNKTTSSVKS 485


Search completed: February 28, 2006, 09:02:02
Job time : 132.62 secs
```

GenCore version 5.1.7

OM protein - protein search, using sw model

Run on: February 28, 2006, 08:56:26 ; Search time 11.9766 Seconds
(without alignments)
556.876 Million cell updates/sec

Title: US-10-717-665A-44_COPY_1_448
Perfect score: 2364
Sequence: 1 MPQPSVSGMDPPFGDAFRSH.....GKRRYFWEYSEQLTPSQQER 448

Scoring table: BLOSUM62
Gapop 10.0 , Gapext 0.5

Searched: 117670 seqs, 14887254 residues

Total number of hits satisfying chosen parameters: 117670

Minimum DB seq length: 0
Maximum DB seq length: 2000000000

Post-processing: Minimum Match 0%
Maximum Match 100%
Listing first 45 summaries

Database : Published_Applications_AA_New:*
1: /cgn2_6/ptodata/2/pubpaa/US08_NEW_PUB.pep:*
2: /cgn2_6/ptodata/2/pubpaa/US06_NEW_PUB.pep:*
3: /cgn2_6/ptodata/2/pubpaa/US07_NEW_PUB.pep:*
4: /cgn2_6/ptodata/2/pubpaa/PCT_NEW_PUB.pep:*
5: /cgn2_6/ptodata/2/pubpaa/US09_NEW_PUB.pep:*
6: /cgn2_6/ptodata/2/pubpaa/US10_NEW_PUB.pep:*
7: /cgn2_6/ptodata/2/pubpaa/US11_NEW_PUB.pep:*
8: /cgn2_6/ptodata/2/pubpaa/US60_NEW_PUB.pep:*

Pred. No. is the number of results predicted by chance to have a score greater than or equal to the score of the result being printed, and is derived by analysis of the total score distribution.

SUMMARIES

Result No.	Score	% Query Match	Length	DB	ID	Description
1	163	6.9	354	7	US-11-189-817-2	Sequence 2, Appli
2	157.5	6.7	651	7	US-11-128-660-1	Sequence 1, Appli
3	155.5	6.6	687	7	US-11-072-512-2651	Sequence 2651, Ap
4	150.5	6.4	740	7	US-11-124-367A-293	Sequence 293, App
5	150.5	6.4	760	7	US-11-124-367A-292	Sequence 292, App
6	141.5	6.0	1229	7	US-11-054-281-107	Sequence 107, App
7	140	5.9	1168	6	US-10-509-422-2	Sequence 2, Appli
8	138.5	5.9	144	6	US-10-821-234-1254	Sequence 1254, Ap
9	138.5	5.9	1017	7	US-11-054-281-108	Sequence 108, App
10	138.5	5.9	1219	7	US-11-054-281-28	Sequence 28, Appl
11	138.5	5.9	1219	7	US-11-054-281-106	Sequence 106, App
12	138.5	5.9	1735	6	US-10-495-083-3	Sequence 3, Appli
13	134	5.7	414	7	US-11-089-551A-28	Sequence 28, Appl
14	132.5	5.6	2801	7	US-11-124-368A-305	Sequence 305, App
15	132.5	5.6	2801	7	US-11-124-367A-433	Sequence 433, App
16	132.5	5.6	2896	7	US-11-124-368A-306	Sequence 306, App
17	132.5	5.6	2896	7	US-11-124-367A-434	Sequence 434, App
18	132.5	5.6	3256	7	US-11-124-368A-304	Sequence 304, App
19	132.5	5.6	3256	7	US-11-124-367A-432	Sequence 432, App
20	131.5	5.6	251	6	US-10-528-031-8	Sequence 8, Appli
21	127	5.4	578	6	US-10-821-234-1039	Sequence 1039, Ap
22	126	5.3	4374	7	US-11-128-572-2	Sequence 2, Appli
23	125	5.3	268	7	US-11-072-512-3158	Sequence 3158, Ap
24	125	5.3	697	6	US-10-821-234-905	Sequence 905, App
25	125	5.3	795	6	US-10-770-726-49	Sequence 49, Appl
26	123.5	5.2	265	6	US-10-799-749-78	Sequence 78, Appl
27	123.5	5.2	770	6	US-10-982-545-15	Sequence 15, Appl
28	123.5	5.2	770	6	US-10-789-273-38	Sequence 38, Appl
29	123	5.2	728	6	US-10-530-340-14	Sequence 14, Appl
30	122.5	5.2	2314	7	US-11-097-728-2	Sequence 2, Appli
31	122.5	5.2	2353	7	US-11-097-728-6	Sequence 6, Appli
32	122	5.2	482	6	US-10-821-234-1413	Sequence 1413, Ap
33	121.5	5.1	743	6	US-10-485-517-351	Sequence 351, App
34	121.5	5.1	877	6	US-10-485-517-200	Sequence 200, App
35	121.5	5.1	902	7	US-11-072-512-2987	Sequence 2987, Ap
36	121	5.1	8746	7	US-11-098-686-10232	Sequence 10232, A
37	120.5	5.1	1596	7	US-11-060-005-4	Sequence 4, Appli
38	120.5	5.1	4384	6	US-10-821-234-1120	Sequence 1120, Ap
39	120	5.1	428	7	US-11-138-642-8	Sequence 8, Appli
40	120	5.1	428	7	US-11-138-882-8	Sequence 8, Appli
41	120	5.1	943	6	US-10-475-204-34	Sequence 34, Appl
42	119.5	5.1	699	7	US-11-138-642-4	Sequence 4, Appli
43	119.5	5.1	699	7	US-11-138-882-4	Sequence 4, Appli
44	119.5	5.1	1405	6	US-10-995-561-529	Sequence 529, App
45	119	5.0	1390	7	US-11-063-343-35	Sequence 35, Appl

ALIGNMENTS

RESULT 1
US-11-189-817-2
; Sequence 2, Application US/11189817
; Publication No. US20060030006A1
; GENERAL INFORMATION:
; APPLICANT: INSTITUT PASTEUR
; APPLICANT: DRUILHE, PIERRE
; TITLE OF INVENTION: IDENTIFICATION OF A CONSERVED REGION OF PLASMODIUM FALCIPARUM
; TITLE OF INVENTION: MSP3 TARGETED BY BIOLOGICALLY ACTIVE ANTIBODIES
; FILE REFERENCE: 275601US0
; CURRENT APPLICATION NUMBER: US/11/189,817
; CURRENT FILING DATE: 2005-07-27
; PRIOR APPLICATION NUMBER: 60/598,062
; PRIOR FILING DATE: 2004-08-03
; NUMBER OF SEQ ID NOS: 14
; SOFTWARE: PatentIn version 3.3
; SEQ ID NO 2
; LENGTH: 354
; TYPE: PRT
; ORGANISM: Plasmodium falciparum
US-11-189-817-2

Query Match 6.9%; Score 163; DB 7; Length 354;
Best Local Similarity 20.9%; Pred. No. 0.00034;
Matches 82; Conservative 61; Mismatches 146; Indels 104; Gaps 13;

```
Qy          51 YQQNPRDNFLSLEDCKDIENLESF-TDVLDNEGALTSNWEQWDTYCEDLTKYTKLTSCDI 109
               |  | |:  |:  :   ||| |:  | :  |:   :      |  | |:| |  | :
Db          32 YNLNLRNAILN--NNSQIENEENVNTTITGND--FSGGEFLWPGYTEEL-KAKKAS---- 82

Qy         110 WGTKEVDYLGLDDFSSPYQDEEVISKTPTLAQLNSEDSQSVSDSLYYPDSLFSVKQNPLP 169
                         :|      | |  ||        :|::   : :|   | ::
Db          83 ----------EDAEKAANDAENASK-------EAEEAAKEAVNLKESDKSYT------- 117

Qy         170 SSFPGKKITSRAAAPVCSSKTLQAEVPLSDCVQKASKPPSSTQIMVKTNMYHNEKVNFHV 229
                     |      |  : | ::  :   |  :|:||  |   |  ||  |  :  | :
Db         118 ------KAKEACTAASKAKKAVETALKAKDDAEKSSKADS---ISTKTKEYAEKAKNAYE 168

Qy         230 ECKDYVKKAKVKINPVQQSRPLLSQIHTDAAKENTCYCGAVAKRQEKKGMEPLQG-HATP 288
               : |:  :||                                || | :|    : : |
Db         169 KAKNAYQKAN--------------------------QAVLKAKEASSYDYILGWEFGG 200

Qy         289 ALPFKETQELLLSPLPQEGPGSLAAGESSSLSASTSVSDSSQKKEEHNYSLFVSDNLGEQ 348
               :|  : :| :|| |                          |:: | |    | | |:
Db         201 GVPEHKKEENMLSHL----------------------YVSSKDKENISKENDDVLDEK 236

Qy         349 PTKCSPEEDEEDEEDVDDEDHDEGFGSEHELSENEEEEEEEEDYEDDKDDDISDTFSEPG 408
                  | :| :||:::::: :|   :| |:||:||||||||: |:: |        :
```

```
Db        237 ----EEEAEETEEEELEEKNEEE---TESEISEDEEEEEEEEEKEEENDKKKEQEKEQSN 289

Qy        409 YENDSVEDLKEVTSISSRKRGKRRYFWEYSEQL 441
                ||  :|::    ||  :    :   | :| :
Db        290 ENNDQKKDMEAQNLISKNQNNNEKNVKEAAESI 322


RESULT 2
US-11-128-660-1
; Sequence 1, Application US/11128660
; Publication No. US20060024324A1
; GENERAL INFORMATION:
;  APPLICANT: Statens Serum Institut
;  TITLE OF INVENTION: Vaccines comprising chimeric malaria proteins derived from Plasmo
;  TITLE OF INVENTION:  falciparum
;  FILE REFERENCE: 15007dk
;  CURRENT APPLICATION NUMBER: US/11/128,660
;  CURRENT FILING DATE:  2005-05-12
;  NUMBER OF SEQ ID NOS: 2
;  SOFTWARE: PatentIn version 3.1
; SEQ ID NO 1
;   LENGTH: 651
;   TYPE: PRT
;   ORGANISM: Plasmodium falciparum
US-11-128-660-1

  Query Match             6.7%;  Score 157.5;  DB 7;  Length 651;
  Best Local Similarity   18.4%;  Pred. No. 0.0017;
  Matches   99;  Conservative   99;  Mismatches  171;  Indels  169;  Gaps   23;

Qy         25 QTLMSTDLLANSSDPDFMYELDRE-MNYQQNPRDNFLSLEDCKDIENLESFTDVL----- 78
              ::|   :|| |||  :    ||:: :: :  |       |:   : | || |   :
Db        130 ESLSDLELLENSSQDN----LDKDTISTEPFPNQKHKDLQQDLNDEPLEPFPTQIHKDYK 185

Qy         79 --------DNEGALTSNWEQWDTYCEDLTKYTKLTSCDIWGTKEVDYLGLDDFSSPYQDE 130
                   |:|       :: | : |:   : :  | :   | :     | |  |    :||
Db        186 EKNLINEEDSEPFPRQEHKKVDNHNEEKNVFHENGSAN--GNQ--GSLKLKSFDEHLKDE 241

Qy        131 EV-----------ISKTPTLAQLNS-EDSQSVSDSLYYPDSLFSVKQNPLPSSFPGKKIT 178
              ::           |   |    || |   :: : ||        |||
Db        242 KIENEPLVHENLSIPNDPIEQILNQPEQETNIQEQLYNEKQNVEEKQN------------ 289

Qy        179 SRAAAPVCSSKTLQAEVPLSDCVQKASKPPSSTQIMVKTNMYHNEK-----VNFHVE--- 230
                      :::|  |      | |::  |:   |   | |     :| ||:
Db        290 --------------SQIPSLDL-----KEPTNEDILPNHNPLENIKQSESEIN-HVQDHA 329

Qy        231 ------------CKDYVKKAKVKINPVQ----------QSRP----------------- 250
                         |::: :::  || :|        | :|
Db        330 LPKENIIDKLDNQKEHIDQSQHNINVLQENNINNHQLEPQEKPNIESFEPKNIDSEIILP 389

Qy        251 -------LLSQIHTDAAKENTCYCGAVAKRQEKKGMEPLQGHATPALPFKETQELLLSPL 303
                    ::  : :     :  :    :: : :: :     | |  :  :||     :|
Db        390 ENVETEEIIDDVPSPKHSNHETFEEETSESEHEEAVSEKNAHET--VEHEETVSQESNPE 447

Qy        304 PQEGPGSLAAGESSSLSASTSV-SDSSQ----------------------------KKEE 334
              :  |:::   :: |: :  | |: |:                             ||||
Db        448 KADNDGNVSQNSNNELNENEFVESEKSEHEARSKAKEASSYDYILGWEFGGGVPEHKKEE 507

Qy        335 HNYS-LFVS----DNLGEQPTKCSPEEDEE----DEEDVDDEDHDEGFGSEHELSENEEE 385
              :  | |:||    :|: ::       |::||    :||:::::: :|   :| |:||:|||
Db        508 NMLSHLYVSSKDKENISKENDDVLDEKEEEAEETEEEELEEKNEEE---TESEISEDEEE 564

Qy        386 EEEEEDYEDD--KDDDISDTFSEPGYENDSVEDLKEVTSISSRKRGKRRYFWEYSEQL 441
              |||||  |::  | :   :  :|    ||  :|::    ||  :    :   | :| :
Db        565 EEEEEKEEENEKKKEQEKEQSNE---NNDQKKDMEAQNLISKNQNNNEKNVKEAAESI 619


RESULT 3
US-11-072-512-2651
; Sequence 2651, Application US/11072512
; Publication No. US20060029945A1
; GENERAL INFORMATION:
;  APPLICANT: ISOGAI, TAKAO
;  APPLICANT:  SUGIYAMA, TOMOYASU
;  APPLICANT:  OTSUKI, TETSUJI
;  APPLICANT:  WAKAMATSU, AI
;  APPLICANT:  SATO, HIROYUKI
;  APPLICANT:  ISHII, SHIZUKO
;  APPLICANT:  YAMAMOTO, JUN-ICHI
;  APPLICANT:  ISONO, YUUKO
;  APPLICANT:  HIO, YURI
;  APPLICANT:  OTSUKA, KAORU
;  APPLICANT:  NAGAI, KEIICHI
;  APPLICANT:  IRIE, RYOTARO
;  APPLICANT:  TAMECHIKA, ICHIRO
;  APPLICANT:  SEKI, NAOHIKO
;  APPLICANT:  YOSHIKAWA, TSUTOMU
;  APPLICANT:  OTSUKA, MOTOYUKI
;  APPLICANT:  NAGAHARI, KENJI
;  APPLICANT:  MASUHO, YASUHIKO
;  TITLE OF INVENTION: Novel full length cDNA
;  FILE REFERENCE: 084335-0191
;  CURRENT APPLICATION NUMBER: US/11/072,512
;  CURRENT FILING DATE:  2005-03-07
;  PRIOR APPLICATION NUMBER: US 60/350,978
;  PRIOR FILING DATE: 2002-01-25
;  PRIOR APPLICATION NUMBER: JP 2001-379298
;  PRIOR FILING DATE: 2001-11-05
;  NUMBER OF SEQ ID NOS: 4096
;  SOFTWARE: PatentIn Ver. 2.1
; SEQ ID NO 2651
;   LENGTH: 687
;   TYPE: PRT
;   ORGANISM: Homo sapiens
US-11-072-512-2651

  Query Match             6.6%;  Score 155.5;  DB 7;  Length 687;
  Best Local Similarity   25.8%;  Pred. No. 0.0024;
  Matches   51;  Conservative   34;  Mismatches   70;  Indels   43;  Gaps    6;

Qy        276 KKGMEPLQGHATPALPFKETQELLLSPLPQEGPGSLAAGESSSLSASTSVSDSSQKKEEH 335
              |||  |  | |  | | |:      : :| :|  :   |:::    |    |   :::|
Db         86 KKGATP--GKALVATPGKKG-----AAIPAKGAKN---GKNAKKEDSDEEEDDDSEEDEE 135

Qy        336 NYSLFVSDNLGEQP--------TKCSPEEDEEDEEDVDDEDHDEGFGSEHE--------- 378
              :      |    :|              | :||:||:|| :|:| ||   || |
Db        136 DDEDEDEDEDEIEPAAMKAAAAAPASEDEDDEDDEDDEDDDDDEEDDSEEEAMETTPAKG 195

Qy        379 ------------LSENEEEEEEEEDYEDDKDDDISDTFSEPGYENDSVED---LKEVTS 422
                          ::|:|:|||::|| :|| |:|  |   |   | :  |:   :||
Db        196 KKAAKVVPVKAKNVAEDEDEEEDDEDEDDDDDEDDEDDDDEDDEEEEEEEEEPVKEAPG 255

Qy        423 ISSRKRGKRRYFWEYSEQ 440
                ::  |::   |  :|
Db        256 KRKKEMAKQKAAPEAKKQ 273


RESULT 4
US-11-124-367A-293
; Sequence 293, Application US/11124367A
; Publication No. US20060024700A1
; GENERAL INFORMATION:
;  APPLICANT: Michele Cargill
;  APPLICANT:  Hongjin Huang
;  TITLE OF INVENTION: Genetic Polymorphisms Associated with
;  TITLE OF INVENTION:  Fibrosis Methods of Detection and Uses Thereof
;  FILE REFERENCE: CL001519.ORD
;  CURRENT APPLICATION NUMBER: US/11/124,367A
;  CURRENT FILING DATE:  2005-05-09
;  PRIOR APPLICATION NUMBER: US 60/568,846
;  PRIOR FILING DATE: 2004-05-07
;  PRIOR APPLICATION NUMBER: US 60/582,609
;  PRIOR FILING DATE: 2004-06-25
;  PRIOR APPLICATION NUMBER: US 60/599,554
;  PRIOR FILING DATE: 2004-08-09
;  NUMBER OF SEQ ID NOS: 34460
```

```
;  SOFTWARE: FastSEQ for Windows Version 4.0
; SEQ ID NO 293
;   LENGTH: 740
;   TYPE: PRT
;   ORGANISM: Homo sapiens
US-11-124-367A-293

  Query Match             6.4%;  Score 150.5;  DB 7;  Length 740;
  Best Local Similarity   23.1%;  Pred. No. 0.0058;
  Matches   83;  Conservative   55;  Mismatches  142;  Indels   79;  Gaps   15;

Qy        113 KEVDYLGLDDFSSPYQDE-----EVISKTPTLAQLNSEDSQSVSDSLY----------YP 157
              ||:|   |||  | |  |     ::|     | :|  :|  |::  :            ||
Db        208 KELDLSELDDPDSAYLQEARLKRKLIRLFGRLCEL--KDCSSLTGRVIEQRIPYRGTRYP 265

Qy        158 DSLFSVKQ---NPLPSSFPGKKITSRAAAPVCSSKTL-----QAEVPLSDCVQKASKPPS 209
              :   :::    | | :||      ||     :  :|     | ::   |  :
Db        266 EVNRRIERLINKPGPDTFPDYGDVLRAVEKAAARHSLGLPRQQLQLMAQDAFRDVG---- 321

Qy        210 STQIMVKTNMYHNEKVNFHVECKDYVKKAKVKINPVQQSRPLLSQIHTDAAKEN------ 263
                | ::   : :   ||      |  :     ::|    | |:|::      :||
Db        322 ---IRLQERRHLDLIYNFGCHLTDDYRPG---VDPA-LSDPVLAR----RLRENRSLAMS 370

Qy        264 ------TCYCGAVAKRQE---KKGMEPLQGHATPALPFKETQELLLSPLPQEGPGSLAA- 313
                   : |     | :|   ||    |||  :: :    :| |  |     |||  :|:
Db        371 RLDEVISKYAMLQDKSEEGERKKRRARLQGTSSHS---ADTPEASLD--SGEGPSGMASQ 425

Qy        314 GESSSLSASTSVSDSSQKKEEHNYSLFVSDNLGEQPTKCSPEEDEEDEEDVDDEDHDEGF 373
              |  |:  | |   |  :  ||                        ||:||:||:  | : :|
Db        426 GCPSASRAETDDEDDEESDEEEE-----------------EEEEEEEEEATDSEEEEDL 467

Qy        374 GSEHELSENEEEEEEEEDYEDDKDDDISDTFSEPGYENDSVEDLKEVTSISSRKRGKRR 432
                |  |::|||:|||:    || | |   |       ::|  |:::  |  :: | |
Db        468 EQMQEGQEDDEEEDEEEEAAAGKDGDKSPMSSLQISNEKNLEPGKQISRSSGEQQNKGR 526


RESULT 5
US-11-124-367A-292
; Sequence 292, Application US/11124367A
; Publication No. US20060024700A1
; GENERAL INFORMATION:
;  APPLICANT: Michele Cargill
;  APPLICANT:  Hongjin Huang
;  TITLE OF INVENTION: Genetic Polymorphisms Associated with
;  TITLE OF INVENTION:  Fibrosis Methods of Detection and Uses Thereof
;  FILE REFERENCE: CL001519.ORD
;  CURRENT APPLICATION NUMBER: US/11/124,367A
;  CURRENT FILING DATE:  2005-05-09
;  PRIOR APPLICATION NUMBER: US 60/568,846
;  PRIOR FILING DATE: 2004-05-07
;  PRIOR APPLICATION NUMBER: US 60/582,609
;  PRIOR FILING DATE: 2004-06-25
;  PRIOR APPLICATION NUMBER: US 60/599,554
;  PRIOR FILING DATE: 2004-08-09
;  NUMBER OF SEQ ID NOS: 34460
;  SOFTWARE: FastSEQ for Windows Version 4.0
; SEQ ID NO 292
;   LENGTH: 760
;   TYPE: PRT
;   ORGANISM: Homo sapiens
US-11-124-367A-292

  Query Match             6.4%;  Score 150.5;  DB 7;  Length 760;
  Best Local Similarity   23.1%;  Pred. No. 0.0059;
  Matches   83;  Conservative   55;  Mismatches  142;  Indels   79;  Gaps   15;

Qy        113 KEVDYLGLDDFSSPYQDE-----EVISKTPTLAQLNSEDSQSVSDSLY----------YP 157
              ||:|   |||  | |  |     ::|     | :|  :|  |::  :            ||
Db        228 KELDLSELDDPDSAYLQEARLKRKLIRLFGRLCEL--KDCSSLTGRVIEQRIPYRGTRYP 285

Qy        158 DSLFSVKQ---NPLPSSFPGKKITSRAAAPVCSSKTL-----QAEVPLSDCVQKASKPPS 209
              :   :::    | | :||      ||     :  :|     | ::   |  :
Db        286 EVNRRIERLINKPGPDTFPDYGDVLRAVEKAAARHSLGLPRQQLQLMAQDAFRDVG---- 341

Qy        210 STQIMVKTNMYHNEKVNFHVECKDYVKKAKVKINPVQQSRPLLSQIHTDAAKEN------ 263
                | ::   : :   ||      |  :     ::|    | |:|::      :||
Db        342 ---IRLQERRHLDLIYNFGCHLTDDYRPG---VDPA-LSDPVLAR----RLRENRSLAMS 390

Qy        264 ------TCYCGAVAKRQE---KKGMEPLQGHATPALPFKETQELLLSPLPQEGPGSLAA- 313
                   : |     | :|   ||    |||  :: :    :| |  |     |||  :|:
Db        391 RLDEVISKYAMLQDKSEEGERKKRRARLQGTSSHS---ADTPEASLD--SGEGPSGMASQ 445

Qy        314 GESSSLSASTSVSDSSQKKEEHNYSLFVSDNLGEQPTKCSPEEDEEDEEDVDDEDHDEGF 373
              |  |:  | |   |  :  ||                        ||:||:||:  | : :|
Db        446 GCPSASRAETDDEDDEESDEEEE-----------------EEEEEEEEEATDSEEEEDL 487

Qy        374 GSEHELSENEEEEEEEEDYEDDKDDDISDTFSEPGYENDSVEDLKEVTSISSRKRGKRR 432
                |  |::|||:|||:    || | |   |       ::|  |:::  |  :: | |
Db        488 EQMQEGQEDDEEEDEEEEAAAGKDGDKSPMSSLQISNEKNLEPGKQISRSSGEQQNKGR 546


RESULT 6
US-11-054-281-107
; Sequence 107, Application US/11054281
; Publication No. US20060013813A1
; GENERAL INFORMATION:
;  APPLICANT: Mezes et al.
;  TITLE OF INVENTION: Proteins and Nucleic Acids Encoding Same
;  FILE REFERENCE: 21402-240CIP
;  CURRENT APPLICATION NUMBER: US/11/054,281
;  CURRENT FILING DATE:  2005-02-08
;  PRIOR APPLICATION NUMBER: 60/261,014
;  PRIOR FILING DATE: 2001-01-11
;  PRIOR APPLICATION NUMBER: 60/261,018
;  PRIOR FILING DATE: 2001-01-11
;  PRIOR APPLICATION NUMBER: 60/318,410
;  PRIOR FILING DATE: 2001-09-10
;  PRIOR APPLICATION NUMBER: 60/261,013
;  PRIOR FILING DATE: 2001-01-11
;  PRIOR APPLICATION NUMBER: 60/261,026
;  PRIOR FILING DATE: 2001-01-11
;  PRIOR APPLICATION NUMBER: 60/261,029
;  PRIOR FILING DATE: 2001-01-11
;  PRIOR APPLICATION NUMBER: 60/313,170
;  PRIOR FILING DATE: 2001-08-17
;  PRIOR APPLICATION NUMBER: 10/044,564
;  PRIOR FILING DATE: 2002-01-11
;  NUMBER OF SEQ ID NOS: 324
;  SOFTWARE: PatentIn Ver. 2.1
; SEQ ID NO 107
;   LENGTH: 1229
;   TYPE: PRT
;   ORGANISM: Rattus norvegicus
US-11-054-281-107

  Query Match             6.0%;  Score 141.5;  DB 7;  Length 1229;
  Best Local Similarity   25.5%;  Pred. No. 0.042;
  Matches   53;  Conservative   40;  Mismatches   50;  Indels   65;  Gaps   13;

Qy        250 PLLSQIHTDAAKENTCYC-------------GAVAK----RQEKKGMEPLQGHATPALPF 292
              | |:    : || :|| |                | ||:       |::|   |     |
Db         45 PTLTMPGRNKAK-STCSCPDLQPNGQDLGESGRVARLGADESEEEG-RSLSNVGDP---- 98

Qy        293 KETQELLLSPLP---------QEGPGSLAAGESSSLSASTSVSDSSQKKEEHNYSLFVSD 343
                |:: ||             | |  :|   : |::   |  :   :::||       :
Db         99 ----EIIKSPSDPKQYRYIKLQNGLQALLISDLSNVEGKTGNATDEEEEEE------EE 148

Qy        344 NLGEQPTKCSPEEDEEDEEDVDDEDHDEGFGSEHELSENEEEE-EEEEDYEDDK--DDDI 400
              ||        ||:||:|:|  ||:|  | |      |: :::||  ::||:::||:  |||:
Db        149 EEGE-------EEEEEEEDDDDDDDEDSG----AEIQDDDEEGFDDEEEFDDDEHDDDDL 197

Qy        401 SDTFSEPGYENDSVEDLKEVTSISSRKR 428
                :       | : :|:|:|   : :||:
Db        198 DN-------EENELEELEE--RVEARKK 216
```

```
RESULT 7
US-10-509-422-2
; Sequence 2, Application US/10509422
; Publication No. US20050244825A1
; GENERAL INFORMATION:
;  APPLICANT: Liou, Simon
;  TITLE OF INVENTION: Human BMP2 Inducible Kinases
;  FILE REFERENCE: 004974.01015
;  CURRENT APPLICATION NUMBER: US/10/509,422
;  CURRENT FILING DATE:  2004-09-24
;  PRIOR APPLICATION NUMBER: PCT/EP03/080825
;  PRIOR FILING DATE: 2003-03-20
;  PRIOR APPLICATION NUMBER: US 60/367,512
;  PRIOR FILING DATE: 2002-03-27
;  PRIOR APPLICATION NUMBER: US 60/406,936
;  PRIOR FILING DATE: 2002-08-30
;  NUMBER OF SEQ ID NOS: 9
;  SOFTWARE: FastSEQ for Windows Version 4.0
; SEQ ID NO 2
;   LENGTH: 1168
;   TYPE: PRT
;   ORGANISM: Homo sapiens
US-10-509-422-2

  Query Match             5.9%;  Score 140;  DB 6;  Length 1168;
  Best Local Similarity   19.3%;  Pred. No. 0.05;
  Matches   90;  Conservative   55;  Mismatches  136;  Indels  186;  Gaps   18;

Qy          2 PQPSVS---GMDPPFGDAFRSHTFSEQTLMSTDLLANSSDPDFMYELDR---------- 47
              |||| |    | | :  ||    | :| |         : |:::
Db        526 PQPSASQYPTMMPQYQQAF----FQQQMLAQHQPSQQQASPEYLTSPQEFSPALVSYTSS 581

Qy         48 ---------EMNYQQNPRDNFLSLEDCKDIENL---ESFTDVLDNEGALTSNWEQWDTYC 95
                      : :|  | :  | |: | : | |    :: ::  |  |          |: :
Db        582 LPAQVGTIMDSSYSANRQVFFQSVADKEAIANFTNQKNISNPPDMSG--------WNPFG 633

Qy         96 EDLTKYTKLTSCDIWGTKEVDYLGLDDFSSPYQDEEVISKTPTLAQLNSEDSQSVSDSLY 155
              ||   ::|||                        :||:: :   | : |  : :: ||
Db        634 ED--NFSKLT----------------------EEELLDREFDLLRSNRLEERASSDK-- 666

Qy        156 YPDSLFSVKQNPLPSSFPGKKITSRAAAPVCSSKTLQAEVPLSDCVQKASKPPSSTQIMV 215
                 ||| :   :|    |                 ||
Db        667 NVDSLSAPHNHPPEDPF-------------------GSVPF------------------ 688

Qy        216 KTNMYHNEKVNFHVECKDYVKKAKVKINPVQQSRPLLSQIHTDAAKENTCYCGAVAKRQE 275
                : |: |                                                |:  |: |
Db        689 ---ISHSGK-------------------------------------------GSPEKKAE 702

Qy        276 KKGMEPLQGHATPALPFKETQELLLSPLPQEGPGSLAAGESSSLSASTSVSDSSQKKEEH 335
                :    | | |    |       ||          | :       |||    ||  :
Db        703 HSSINQENGTANPIKNGK------TSP----------ASKDQRTGKKTSVQGQVQKGNDE 746

Qy        336 NYSLFVSDNLGEQPTKCSPEEDEEDEEDV---------DDEDHDEGFGSEHELSENEEEE 386
                : | | ||      |:  | ||:|:|:|:|          ||:   | |    | ::|:||
Db        747 SESDFESD----PPSPKSSEEEEQDDEEVLQGEQGDFNDDDTEPENLGHRPLLMDSEDEE 802

Qy        387 EEEE-----DYEDDKD--DDISDTFSEPGYENDSVEDLKEVTSISSR 426
                |||:     |||  |    |:|  : :    :   :||  :   |::
Db        803 EEEKHSSDSDYEQAKAKYSDMSSVYRDRS-GSGPTQDLNTILLTSAQ 848


RESULT 8
US-10-821-234-1254
; Sequence 1254, Application US/10821234
; Publication No. US20050255114A1
; GENERAL INFORMATION:
;  APPLICANT: Labat, Ivan
;  APPLICANT:  Stache-Crain, Birgit
;  APPLICANT:  Andarmani, Susan
;  APPLICANT:  Tang, Y. Tom
;  TITLE OF INVENTION: Methods for Diagnosis and Treatment of Preeclampsia
;  FILE REFERENCE: 821A
;  CURRENT APPLICATION NUMBER: US/10/821,234
;  CURRENT FILING DATE:  2004-04-07
;  PRIOR APPLICATION NUMBER: US 60/462,047
;  PRIOR FILING DATE: 2003-04-07
;  NUMBER OF SEQ ID NOS: 1704
;  SOFTWARE: pt_SEQ_genes Version 1.0
; SEQ ID NO 1254
;   LENGTH: 144
;   TYPE: PRT
;   ORGANISM: Homo sapiens
US-10-821-234-1254

  Query Match             5.9%;  Score 138.5;  DB 6;  Length 144;
  Best Local Similarity   30.5%;  Pred. No. 0.0053;
  Matches   32;  Conservative   20;  Mismatches   28;  Indels   25;  Gaps    2;

Qy        353 SPEEDEEDEEDVDDEDHDEGFGSEHE----------------------LSENEEEEEEEE 390
              | :||:||:|| :|:| ||    || |                      ::|:|:|||:|
Db         40 SEDEDDDEDDEDDEDDDDDEEDDSEEEAMETTPAKGKKAAKVVPVKAKNVAEDEDEEEDDE 99

Qy        391 DYEDDKDDDISDTFSEPGYENDSVED---LKEVTSISSRKRGKRR 432
              | :|| |:|  |    |    | :  |:   :||      ::  |::
Db        100 DEDDDDDEDDEDDDDEDDEEEEEEEEEEPVKEAPGKRKKEMAKQK 144


RESULT 9
US-11-054-281-108
; Sequence 108, Application US/11054281
; Publication No. US20060013813A1
; GENERAL INFORMATION:
;  APPLICANT: Mezes et al.
;  TITLE OF INVENTION: Proteins and Nucleic Acids Encoding Same
;  FILE REFERENCE: 21402-240CIP
;  CURRENT APPLICATION NUMBER: US/11/054,281
;  CURRENT FILING DATE:  2005-02-08
;  PRIOR APPLICATION NUMBER: 60/261,014
;  PRIOR FILING DATE: 2001-01-11
;  PRIOR APPLICATION NUMBER: 60/261,018
;  PRIOR FILING DATE: 2001-01-11
;  PRIOR APPLICATION NUMBER: 60/318,410
;  PRIOR FILING DATE: 2001-09-10
;  PRIOR APPLICATION NUMBER: 60/261,013
;  PRIOR FILING DATE: 2001-01-11
;  PRIOR APPLICATION NUMBER: 60/261,026
;  PRIOR FILING DATE: 2001-01-11
;  PRIOR APPLICATION NUMBER: 60/261,029
;  PRIOR FILING DATE: 2001-01-11
;  PRIOR APPLICATION NUMBER: 60/313,170
;  PRIOR FILING DATE: 2001-08-17
;  PRIOR APPLICATION NUMBER: 10/044,564
;  PRIOR FILING DATE: 2002-01-11
;  NUMBER OF SEQ ID NOS: 324
;  SOFTWARE: PatentIn Ver. 2.1
; SEQ ID NO 108
;   LENGTH: 1017
;   TYPE: PRT
;   ORGANISM: Homo sapiens
US-11-054-281-108

  Query Match             5.9%;  Score 138.5;  DB 7;  Length 1017;
  Best Local Similarity   29.0%;  Pred. No. 0.054;
  Matches   51;  Conservative   35;  Mismatches   61;  Indels   29;  Gaps    9;

Qy        250 PLLSQIHTDAAKENTCYCGAVAKRQEKKGMEPLQGHATPALPFKETQELLLSPLPQEG-P 308
              |:|:    : || :|| |  :    :  | :  :       |   |::|         ||
Db         46 PILAMPGRNKAK-STCSCPDL----QPNGQDLGENSRVARLGADESEE--------EGRR 92

Qy        309 GSLA-AGESSSLSASTSVSDSSQ----KKEEHNYSLFVSDNLGEQPTKCSPEEDEEDEED 363
              |||: ||:    :    | ||  |    |  :    :| :|| |      |      |:|:||:
Db         93 GSLSNAGDPEIVK---SPSDPKQYRYIKLQNGLQALLISD-LSNMEGKTGNTTDDEEEEE 148

Qy        364 VDDEDHDEGFGSEHELSENEEEEEEEDYEDDKDDDISDTFSEPGYENDSVEDLKE 419
              |::|: |:    |  |: :::||  ::||  ||: ||  ||       |:: :|:|:|
Db        149 VEEEEEDDDEDSGAEIEDDDEEGFDDEDEFDDEHDDDLDT------EDNELEELEE 198
```

```
RESULT 10
US-11-054-281-28
; Sequence 28, Application US/11054281
; Publication No. US20060013813A1
; GENERAL INFORMATION:
;  APPLICANT: Mezes et al.
;  TITLE OF INVENTION: Proteins and Nucleic Acids Encoding Same
;  FILE REFERENCE: 21402-240CIP
;  CURRENT APPLICATION NUMBER: US/11/054,281
;  CURRENT FILING DATE:  2005-02-08
;  PRIOR APPLICATION NUMBER: 60/261,014
;  PRIOR FILING DATE: 2001-01-11
;  PRIOR APPLICATION NUMBER: 60/261,018
;  PRIOR FILING DATE: 2001-01-11
;  PRIOR APPLICATION NUMBER: 60/318,410
;  PRIOR FILING DATE: 2001-09-10
;  PRIOR APPLICATION NUMBER: 60/261,013
;  PRIOR FILING DATE: 2001-01-11
;  PRIOR APPLICATION NUMBER: 60/261,026
;  PRIOR FILING DATE: 2001-01-11
;  PRIOR APPLICATION NUMBER: 60/261,029
;  PRIOR FILING DATE: 2001-01-11
;  PRIOR APPLICATION NUMBER: 60/313,170
;  PRIOR FILING DATE: 2001-08-17
;  PRIOR APPLICATION NUMBER: 10/044,564
;  PRIOR FILING DATE: 2002-01-11
;  NUMBER OF SEQ ID NOS: 324
;  SOFTWARE: PatentIn Ver. 2.1
; SEQ ID NO 28
;   LENGTH: 1219
;   TYPE: PRT
;   ORGANISM: Homo sapiens
US-11-054-281-28

  Query Match             5.9%;  Score 138.5;  DB 7;  Length 1219;
  Best Local Similarity   29.0%;  Pred. No. 0.067;
  Matches   51;  Conservative   35;  Mismatches   61;  Indels   29;  Gaps    9;

Qy        250 PLLSQIHTDAAKENTCYCGAVAKRQEKKGMEPLQGHATPALPFKETQELLLSPLPQEG-P 308
              |:|:    : || :|| | :    :  | :  :       |   |::|         ||
Db         46 PILAMPGRNKAK-STCSCPDL----QPNGQDLGENSRVARLGADESEE--------EGRR 92

Qy        309 GSLA-AGESSSLSASTSVSDSSQ----KKEEHNYSLFVSDNLGEQPTKCSPEEDEEDEED 363
              |||: ||:   :    | ||  |    | :    :| :|| |      |      |:|:||:
Db         93 GSLSNAGDPEIVK---SPSDPKQYRYIKLQNGLQALLISD-LSNMEGKTGNTTDDEEEEE 148

Qy        364 VDDEDHDEGFGSEHELSENEEEEEEEEDYEDDKDDDISDTFSEPGYENDSVEDLKE 419
              |::|: |:   |  |: :::||  ::||  ||: ||  ||      |:: :|:|:|
Db        149 VEEEEEDDDEDSGAEIEDDDEEGFDDEDEFDDEHDDDLDT------EDNELEELEE 198


RESULT 11
US-11-054-281-106
; Sequence 106, Application US/11054281
; Publication No. US20060013813A1
; GENERAL INFORMATION:
;  APPLICANT: Mezes et al.
;  TITLE OF INVENTION: Proteins and Nucleic Acids Encoding Same
;  FILE REFERENCE: 21402-240CIP
;  CURRENT APPLICATION NUMBER: US/11/054,281
;  CURRENT FILING DATE:  2005-02-08
;  PRIOR APPLICATION NUMBER: 60/261,014
;  PRIOR FILING DATE: 2001-01-11
;  PRIOR APPLICATION NUMBER: 60/261,018
;  PRIOR FILING DATE: 2001-01-11
;  PRIOR APPLICATION NUMBER: 60/318,410
;  PRIOR FILING DATE: 2001-09-10
;  PRIOR APPLICATION NUMBER: 60/261,013
;  PRIOR FILING DATE: 2001-01-11
;  PRIOR APPLICATION NUMBER: 60/261,026
;  PRIOR FILING DATE: 2001-01-11
;  PRIOR APPLICATION NUMBER: 60/261,029
;  PRIOR FILING DATE: 2001-01-11
;  PRIOR APPLICATION NUMBER: 60/313,170
;  PRIOR FILING DATE: 2001-08-17
;  PRIOR APPLICATION NUMBER: 10/044,564
;  PRIOR FILING DATE: 2002-01-11
;  NUMBER OF SEQ ID NOS: 324
;  SOFTWARE: PatentIn Ver. 2.1
; SEQ ID NO 106
;   LENGTH: 1219
;   TYPE: PRT
;   ORGANISM: Homo sapiens
US-11-054-281-106

  Query Match             5.9%;  Score 138.5;  DB 7;  Length 1219;
  Best Local Similarity   29.0%;  Pred. No. 0.067;
  Matches   51;  Conservative   35;  Mismatches   61;  Indels   29;  Gaps    9;

Qy        250 PLLSQIHTDAAKENTCYCGAVAKRQEKKGMEPLQGHATPALPFKETQELLLSPLPQEG-P 308
              |:|:    : || :|| | :    :  | :  :       |   |::|         ||
Db         46 PILAMPGRNKAK-STCSCPDL----QPNGQDLGENSRVARLGADESEE--------EGRR 92

Qy        309 GSLA-AGESSSLSASTSVSDSSQ----KKEEHNYSLFVSDNLGEQPTKCSPEEDEEDEED 363
              |||: ||:   :    | ||  |    | :    :| :|| |      |      |:|:||:
Db         93 GSLSNAGDPEIVK---SPSDPKQYRYIKLQNGLQALLISD-LSNMEGKTGNTTDDEEEEE 148

Qy        364 VDDEDHDEGFGSEHELSENEEEEEEEEDYEDDKDDDISDTFSEPGYENDSVEDLKE 419
              |::|: |:   |  |: :::||  ::||  ||: ||  ||      |:: :|:|:|
Db        149 VEEEEEDDDEDSGAEIEDDDEEGFDDEDEFDDEHDDDLDT------EDNELEELEE 198


RESULT 12
US-10-495-083-3
; Sequence 3, Application US/10495083
; Publication No. US20060008800A1
; GENERAL INFORMATION:
;  APPLICANT: Max-Planck-Gesellschaft zur Forderung der Wissenschaften e.V.
;  APPLICANT:  ROSENMUND, Christian
;  APPLICANT:  BROSE, Nils
;  APPLICANT:  RHEE, Jeong-Seop
;  APPLICANT:  BETZ, Andrea
;  APPLICANT:  RETTIG, Jens
;  APPLICANT:  ASHERY, Uri
;  APPLICANT:  JUNGE, Harald
;  TITLE OF INVENTION: UNC-13 IN THE MODULATION OF NEUROTRANSMISSION AND SECRETION EVENTS
;  FILE REFERENCE: 009848-0309442
;  CURRENT APPLICATION NUMBER: US/10/495,083
;  CURRENT FILING DATE:  2004-05-05
;  PRIOR APPLICATION NUMBER: PCT/EP02/12072
;  PRIOR FILING DATE: 2002-10-29
;  PRIOR APPLICATION NUMBER: EP 01 12 6235.9
;  PRIOR FILING DATE: 2001-11-05
;  NUMBER OF SEQ ID NOS: 12
;  SOFTWARE: PatentIn version 3.1
; SEQ ID NO 3
;   LENGTH: 1735
;   TYPE: PRT
;   ORGANISM: Rattusnorvegicus
US-10-495-083-3

  Query Match             5.9%;  Score 138.5;  DB 6;  Length 1735;
  Best Local Similarity   22.1%;  Pred. No. 0.1;
  Matches  103;  Conservative   59;  Mismatches  141;  Indels  163;  Gaps   25;

Qy         21 TFSEQTLMSTDLLANSSDP----DFMYELDREMNYQQNPRDNFLSLEDCKDIENLESFTD 76
              |   | : || :    | |    |||:|::|          |  |::|
Db         25 TLKVQNVKSTTIAVRGSQPSWEQDFMFEINR--------LDLGLTVE------------ 63

Qy         77 VLDNEGALTSNWEQWDTYC------------------EDLTKYTK--LTSCDIWGTKEV 115
               : |:| :      |||                  | ||  ::  :   :| |||:
Db         64 -VWNKGLI------WDTMVGTVWIPLRTIRQSNEEGPGEWLTLDSQAIMADSEICGTKDP 116

Qy        116 DY--LGLD-DFSSPYQ--DEEVISKTPTLAQLNS---EDSQSVSDSLYYPDSLFSVKQNP 167
```

```
Db        117 TFHRILLDAHFELPLDIPEEEARYWAKKLEQLNAMRDQDEYSFQDQ----------QDKP 166

Qy        168 LPSSFPGKKITSRAAAPVCSSKTLQAEVPLSDCVQ-------KASKPPSSTQIMVKTNMY 220

Db        167 LP------------------------VPSSQCCNWNYFGWGEQNDDPDSA-VDDRDSDY 200

Qy        221 HNEKVN-----FHVECKDYVKKAKVKINPVQQSRPLLSQIHTDAAKENTCYCGAVAKRQE 275

Db        201 RSETSNSIPPPYYTTSQPNASVHQYSVRP----PPLGSR---------ESYSDSMHSYEE 247

Qy        276 KKGMEPLQGHATPALPFKETQELLLSPLPQEGPGSLAAGESSSLSASTSVSDSSQKKEEH 335

Db        248 ---------------FSEPR--ALSP-----TGSSRYASSGELSQGSSQLSEDFDPDEH 284

Qy        336 NYSLFVSDNLGEQPT-----------KCSP--EEDEEDEEDVDDEDHDEGFGSEHELSE- 381

Db        285 SLQGSELDDERDRDSYHSCHSSVSYHKDSPRWDQDEEDLEDLEDLE-DEELPEEEELEEE 343

Qy        382 --NEEEEEEEEDYEDDKDDDISDTF-SEPGYENDSVEDLKEVTSIS 424

Db        344 ELEEEEELEEEELELEEEEEVPDDLASYTQQEDTTVAEPKEFKRIS 389


RESULT 13
US-11-089-551A-28
; Sequence 28, Application US/11089551A
; Publication No. US20050266242A1
; GENERAL INFORMATION:
;  APPLICANT: Lindquist et al.
;  TITLE OF INVENTION: ELECTRICAL CONDUCTORS AND DEVICES FROM PRION-LIKE PROTEINS
;  FILE REFERENCE: 30554/40025A
;  CURRENT APPLICATION NUMBER: US/11/089,551A
;  CURRENT FILING DATE:  2005-03-24
;  PRIOR APPLICATION NUMBER: US 60/559,286
;  PRIOR FILING DATE: 2004-03-31
;  NUMBER OF SEQ ID NOS: 65
;  SOFTWARE: PatentIn Ver. 2.0
; SEQ ID NO 28
;   LENGTH: 414
;   TYPE: PRT
;   ORGANISM: Saccharomyces cerevisiae
US-11-089-551A-28

  Query Match             5.7%;  Score 134;  DB 7;  Length 414;
  Best Local Similarity   30.5%;  Pred. No. 0.037;
  Matches   36;  Conservative   25;  Mismatches   53;  Indels    4;  Gaps    2;

Qy        310 SLAAGESSSLSASTSVSDSSQKKEEHNYSLFVSDNLGEQPTKCSPEEDEEDEEDVDDEDH 369

Db         29 SSSSSESSSSSSSSSESESESESESESSS---SSSSSDSESSSSSSSDSESEAETKKEES 85

Qy        370 DEGFGSEHELSENEEEEEEEEDYEDDKDDDISDTFSEPGYENDSVEDLKEVTSISSRK 427

Db         86 KDSSSSSSDSSSDEEEEEEKEETKKEESKESSSSDSSSSSSSDS-ESEKEESNDKKRK 142


RESULT 14
US-11-124-368A-305
; Sequence 305, Application US/11124368A
; Publication No. US20050287559A1
; GENERAL INFORMATION:
;  APPLICANT: Michele Cargill
;  APPLICANT:  James J. Devlin
;  APPLICANT:  May Luke
;  TITLE OF INVENTION: Genetic Polymorphisms Associated with
;  TITLE OF INVENTION:  Vascular Diseases, Methods of Detection and Uses Thereof
;  FILE REFERENCE: CL001524
;  CURRENT APPLICATION NUMBER: US/11/124,368A
;  CURRENT FILING DATE:  2005-05-09
;  PRIOR APPLICATION NUMBER: US 60/568,845
;  PRIOR FILING DATE: 2004-05-07
;  PRIOR APPLICATION NUMBER: US 60/625,936
;  PRIOR FILING DATE: 2004-11-09
;  NUMBER OF SEQ ID NOS: 21112
;  SOFTWARE: FastSEQ for Windows Version 4.0
; SEQ ID NO 305
;   LENGTH: 2801
;   TYPE: PRT
;   ORGANISM: Homo sapiens
US-11-124-368A-305

  Query Match             5.6%;  Score 132.5;  DB 7;  Length 2801;
  Best Local Similarity   20.8%;  Pred. No. 0.46;
  Matches  110;  Conservative   65;  Mismatches  184;  Indels  169;  Gaps   21;

Qy          2 PQPSVSGMDPPFGDAFRSHTFSEQTLMSTDLLANSSDPDFMYELDREMNYQQNPRDNFLS 61

Db       1935 PKPAVS-------DEKNINTFVETPVQKLDLLGN---------LPGSKRQPQTPKEKAEA 1978

Qy         62 LEDCKDIENL--------ESFTDVLDNEGALTS-NWEQWDT----------------YCE 96

Db       1979 LEDLVGFKELFQTPGHTEESMTDDKITEVSCKSPQPESFKTSRSSKQRLKIPLVKVDMKE 2038

Qy         97 DLTKYTKLT----------------SCDIWGTKEVDYLGLDDFSSPYQD-EEVISKTPTL 139

Db       2039 EPLAVSKLTRTSGETTQTHTEPTGDSKSIKAFKE----------SPKQILDPAASVTGSR 2088

Qy        140 AQLNS--EDSQSVSDSLYYPDSLFSV-----------KQNPLPSSFPGKKITSRAAAPVC 186

Db       2089 RQLRTRKEKARALEDLVDFKE-LFSAPGHTEESMTIDKNTKIPCKSPPPELTDTAT---- 2143

Qy        187 SSKTLQAEVPLSDCVQKASKPPSSTQIMVKTNMYHNEKVNFHVECKDYVKKAKVKINPVQ 246

Db       2144 STKRCPKTRPRKEVKEELSAVERLTQTSGQSTHTHKEPASGDEGIKVLKQRAKKKPNPVE 2203

Qy        247 QS----RPLLSQIHTDAAKENTCYCGAVAKRQEKKGMEPLQGHATPALPFKETQELLLSP 302

Db       2204 EEPSRRRPR------------------APKEKAQPLEDLAGFTELSETSGHTQE----- 2239

Qy        303 LPQEGPGSLAAGESSSL---SASTSVSDSSQKKEEHNYSLFVSDNLGEQP-----TKCSP 354

Db       2240 -------SLTAGKATKIPCESPPLEVVDTTASTKRHLRTRVQKVQVKEEPSAVKFTQTSG 2292

Qy        355 EEDEEDEEDVDDEDHDEGF-----------------GSEHELSENEEEEEEEEDYEDDK 396

Db       2293 ETTDADKEPAGE---DKGIKALKESAKQTPAPAASVTGSRRRPRAPRESAQAIEDLAGFK 2349

Qy        397 DDDISDTFSEPGYENDSVEDLK------------EVTSISSRKRGKRR 432

Db       2350 D-------PAAGHTEESMTDDKTTKIPCKSSPELEDTATSSKRRPRTR 2390


RESULT 15
US-11-124-367A-433
; Sequence 433, Application US/11124367A
; Publication No. US20060024700A1
; GENERAL INFORMATION:
;  APPLICANT: Michele Cargill
;  APPLICANT:  Hongjin Huang
;  TITLE OF INVENTION: Genetic Polymorphisms Associated with
;  TITLE OF INVENTION:  Fibrosis Methods of Detection and Uses Thereof
;  FILE REFERENCE: CL001519.ORD
;  CURRENT APPLICATION NUMBER: US/11/124,367A
;  CURRENT FILING DATE:  2005-05-09
;  PRIOR APPLICATION NUMBER: US 60/568,846
;  PRIOR FILING DATE: 2004-05-07
;  PRIOR APPLICATION NUMBER: US 60/582,609
;  PRIOR FILING DATE: 2004-06-25
;  PRIOR APPLICATION NUMBER: US 60/599,554
;  PRIOR FILING DATE: 2004-08-09
;  NUMBER OF SEQ ID NOS: 34460
;  SOFTWARE: FastSEQ for Windows Version 4.0
; SEQ ID NO 433
;   LENGTH: 2801
;   TYPE: PRT
```

; ORGANISM: Homo sapiens
US-11-124-367A-433

Query Match 5.6%; Score 132.5; DB 7; Length 2801;
Best Local Similarity 20.8%; Pred. No. 0.46;
Matches 110; Conservative 65; Mismatches 184; Indels 169; Gaps 21;

```
Qy          2 PQPSVSGMDPPFGDAFRSHTFSEQTLMSTDLLANSSDPDFMYELDREMNYQQNPRDNFLS 61
              |:|:||       |    :|| |  :   ||| |         |        | |::   :
Db       1935 PKPAVS-------DEKNINTFVETPVQKLDLLGN---------LPGSKRQPQTPKEKAEA 1978

Qy         62 LEDCKDIENL--------ESFTDVLDNEGALTS-NWEQWDT---------------YCE 96
              |||    : |        || ||    | :  |   | : |                |
Db       1979 LEDLVGFKELFQTPGHTEESMTDDKITEVSCKSPQPESFKTSRSSKQRLKIPLVKVDMKE 2038

Qy         97 DLTKYTKLT---------------SCDIWGTKEVDYLGLDDFSSPYQD-EEVISKTPTL 139
              :    :|||                |  |   ||          || |  :   | | :
Db       2039 EPLAVSKLTRTSGETTQTHTEPTGDSKSIKAFKE----------SPKQILDPAASVTGSR 2088

Qy        140 AQLNS--EDSQSVSDSLYYPDSLFSV----------KQNPLPSSFPGKKITSRAAAPVC 186
               || :  | :::: | : : : |||            |   :|   |  ::|  |
Db       2089 RQLRTRKEKARALEDLVDFKE-LFSAPGHTEESMTIDKNTKIPCKSPPPELTDTAT---- 2143

Qy        187 SSKTLQAEVPLSDCVQKASKPPSSTQIMVKTNMYHNEKVNFHVECKDYVKKAKVKINPVQ 246
              |:|      |  :  :: |     ||   ::   | |  :      |   ::|| | |||:
Db       2144 STKRCPKTRPRKEVKEELSAVERLTQTSGQSTHTHKEPASGDEGIKVLKQRAKKKPNPVE 2203

Qy        247 QS----RPLLSQIHTDAAKENTCYCGAVAKRQEKKGMEPLQGHATPALPFKETQELLLSP 302
              :     ||                    | ::: : :| | |    :     |||
Db       2204 EEPSRRRPR-----------------APKEKAQPLEDLAGFTELSETSGHTQE----- 2239

Qy        303 LPQEGPGSLAAGESSSL---SASTSVSDSSQKKEEHNYSLFVSDNLGEQP-----TKCSP 354
                 || ||::: :   |    | |::   : |  :      : |:|     |: |
Db       2240 -------SLTAGKATKIPCESPPLEVVDTTASTKRHLRTRVQKVQVKEEPSAVKFTQTSG 2292

Qy        355 EEDEEDEEDVDDEDHDEGF-----------------GSEHELSENEEEEEEEEDYEDDK 396
              |  : |:|   :   |:|                 ||      |  : ||   |
Db       2293 ETTDADKEPAGE---DKGIKALKESAKQTPAPAASVTGSRRRPRAPRESAQAIEDLAGFK 2349

Qy        397 DDDISDTFSEPGYENDSVEDLK-----------EVTSISSRKRGKRR 432
              |          |:  :|: | |            | |: ||::| : |
Db       2350 D-------PAAGHTEESMTDDKTTKIPCKSSPELEDTATSSKRRPRTR 2390
```

Search completed: February 28, 2006, 09:02:40
Job time : 13.9766 secs

This Page Blank (uspto)

GenCore version 5.1.7

OM protein - protein search, using sw model

Run on: February 28, 2006, 08:45:35 ; Search time 27.3216 Seconds
(without alignments)
1577.691 Million cell updates/sec

Title: US-10-717-665A-44_COPY_1_448
Perfect score: 2364
Sequence: 1 MPQPSVSGMDPPFGDAFRSH.....GKRRYFWEYSEQLTPSQQER 448

Scoring table: BLOSUM62
Gapop 10.0 , Gapext 0.5

Searched: 283416 seqs, 96216763 residues

Total number of hits satisfying chosen parameters: 283416

Minimum DB seq length: 0
Maximum DB seq length: 2000000000

Post-processing: Minimum Match 0%
Maximum Match 100%
Listing first 45 summaries

Database : PIR_80:*
1: pir1:*
2: pir2:*
3: pir3:*
4: pir4:*

Pred. No. is the number of results predicted by chance to have a score greater than or equal to the score of the result being printed, and is derived by analysis of the total score distribution.

SUMMARIES

Result No.	Score	% Query Match	Length	DB	ID	Description
1	188.5	8.0	707	1	DNMS	nucleolin - mouse
2	185.5	7.8	712	2	JH0148	nucleolin - rat
3	178.5	7.6	2761	2	T29285	hypothetical prote
4	165	7.0	678	2	A54514	glutamic acid-rich
5	165	7.0	707	2	A35804	nucleolin - human
6	162	6.9	1187	2	T46637	transcription fact
7	161	6.8	831	2	S39835	hypothetical prote
8	160.5	6.8	1974	2	T16703	hypothetical prote
9	159.5	6.7	1128	2	G86266	hypothetical prote
10	157	6.6	562	2	S38149	SIS2 protein - yea
11	153	6.5	792	2	T42963	hypothetical prote
12	152.5	6.5	500	2	S55785	nucleolar protein
13	152.5	6.5	896	2	T24169	hypothetical prote
14	152.5	6.5	906	2	T24166	hypothetical prote
15	152	6.4	1038	2	S52522	hypothetical prote
16	152	6.4	1182	2	T30189	myelin transcripti
17	151.5	6.4	740	2	T03847	Fas-binding protei
18	151.5	6.4	1188	2	T46608	zinc finger protei
19	150.5	6.4	699	2	T01029	hypothetical prote
20	150.5	6.4	713	2	A27441	nucleolin - Chines
21	150.5	6.4	797	2	A36811	hypothetical prote
22	149	6.3	1063	2	A40253	acidic nuclear pro
23	149	6.3	1262	2	T33074	hypothetical prote
24	148.5	6.3	1089	2	S48244	NMD2 protein - yea
25	148.5	6.3	1094	2	S49313	protein kinase - s
26	148.5	6.3	1300	2	T03166	probable immediate
27	148	6.3	972	2	T49773	related to actin-i
28	147	6.2	3724	2	T18427	hypothetical prote
29	146.5	6.2	727	2	JC5113	ribosomal transcri
30	146.5	6.2	764	2	JC5112	ribosomal transcri
31	146.5	6.2	1804	2	T34518	nestin - golden ha
32	146	6.2	694	1	DNCHNL	nucleolin - chicke
33	146	6.2	701	2	S17196	transcription fact
34	146	6.2	913	2	T52485	neurofilament prot
35	146	6.2	933	2	S41539	fibrinogen-binding
36	145.5	6.2	407	1	EDBEQ3	immediate-early pr
37	145.5	6.2	553	2	T15094	hypothetical prote
38	145	6.1	990	2	T38274	probable transcrip
39	145	6.1	1085	2	S55352	IFH1 protein - yea
40	144.5	6.1	409	2	E86336	hypothetical prote
41	144.5	6.1	1192	2	A71623	probable secreted
42	143	6.0	1100	2	T30967	transcription acti
43	143	6.0	1115	2	B84476	probable TPR repea
44	142.5	6.0	279	2	S30766	ASF1 protein - yea
45	142	6.0	482	2	S37845	transcription init

ALIGNMENTS

RESULT 1
DNMS
nucleolin - mouse
C;Species: Mus musculus (house mouse)
C;Date: 30-Sep-1992 #sequence_revision 30-Sep-1992 #text_change 09-Jul-2004
C;Accession: A29958; A40769; A56240; I84688
R;Bourbon, H.M.; Lapeyre, B.; Amalric, F.
J. Mol. Biol. 200, 627-638, 1988
A;Title: Structure of the mouse nucleolin gene. The complete sequence reveals that each I
A;Reference number: A29958; MUID:88316930; PMID:3137346
A;Accession: A29958
A;Molecule type: DNA
A;Residues: 1-707 <BOU>
A;Cross-references: UNIPROT:P09405; UNIPARC:UPI00000009C3; GB:X07699; NID:g53453; PIDN:C
R;Pasternack, M.S.; Bleier, K.J.; McInerney, T.N.
J. Biol. Chem. 266, 14703-14708, 1991
A;Title: Granzyme A binding to target cell proteins. Granzyme A binds to and cleaves nucl
A;Reference number: A40769; MUID:91317840; PMID:1860869
A;Accession: A40769
A;Molecule type: protein
A;Residues: 2-20,'X',22-24 <PAS>
A;Cross-references: UNIPARC:UPI00001739CF
R;Yang, T.H.; Tsai, W.H.; Lee, Y.M.; Lei, H.Y.; Lai, M.Y.; Chen, D.S.; Yeh, N.H.; Lee, S.
Mol. Cell. Biol. 14, 6068-6074, 1994
A;Title: Purification and characterization of nucleolin and its identification as a trans
A;Reference number: A56240; MUID:94344117; PMID:8065340
A;Accession: A56240
A;Molecule type: protein
A;Residues: 2-19;558-567 <YAN>
A;Cross-references: UNIPARC:UPI00001739D0; UNIPARC:UPI00001739D1
R;Bourbon, H.
Gene 68, 73-84, 1988
A;Title: Sequence and structure of the nucleolin promoter in rodents: Characterization of
A;Reference number: I48118; MUID:89121496; PMID:2906027
A;Accession: I84688
A;Status: preliminary; translated from GB/EMBL/DDBJ
A;Molecule type: DNA
A;Residues: 1-44 <RES>
A;Cross-references: UNIPARC:UPI000016CF68; GB:M22089; NID:g200111; PIDN:AAA39841.1; PID:g
C;Comment: This housekeeping protein is involved in the synthesis, packaging, and maturat
C;Genetics:
A;Introns: 6/3; 45/3; 211/1; 273/1; 302/1; 349/2; 391/1; 432/3; 484/1; 525/2; 566/1; 608/
C;Superfamily: nucleolin; ribonucleoprotein repeat homology
C;Keywords: DNA binding; duplication; nucleus; phosphoprotein; RNA binding; transcriptior
F;310-375/Domain: ribonucleoprotein repeat homology <RRM1>
F;311-316/Region: RNA-binding RNP2 motif
F;349-356/Region: RNA-binding RNP1 motif
F;396-458/Domain: ribonucleoprotein repeat homology <RRM2>
F;397-402/Region: RNA-binding RNP2 motif
F;431-438/Region: RNA-binding RNP1 motif
F;488-551/Domain: ribonucleoprotein repeat homology <RRM3>
F;489-494/Region: RNA-binding RNP2 motif

```
F;524-531/Region: RNA-binding RNP1 motif
F;570-634/Domain: ribonucleoprotein repeat homology <RRM4>
F;571-576/Region: RNA-binding RNP2 motif
F;607-614/Region: RNA-binding RNP1 motif

  Query Match             8.0%;  Score 188.5;  DB 1;  Length 707;
  Best Local Similarity   23.5%;  Pred. No. 0.002;
  Matches   84;  Conservative   40;  Mismatches  113;  Indels  121;  Gaps   12;

Qy        137 PTLAQLNSEDSQSVSDSLYYPDSLFSVKQNPLPSSFPGKKITSRAAAPVCSSKTLQAEVP 196
              |   : :||| :    |     |      ::  :|     ||| |:  |  |  |:| :| ||
Db         21 PKEVEEDSEDEEMSEDE----DDSSGEEEVVIPQK-KGKKATTTPAKKVVVSQTKKAAVP 75

Qy        197 LSDCVQKASKPPSSTQIMVKTNMYHNEKVNFHVECKDYVKKAKVKINPVQQSRPLLSQIH 256
                :||:  |                        |     ||  | | :
Db         76 TP--AKKAAVTPGK--------------------KAVATPAKKNITPAK---------- 102

Qy        257 TDAAKENTCYCGAVAKRQEKKGMEPLQGHATPALPFKETQELLLSPLP-QEGPGSLAAGE 315
                      |     |||                 |   | | | ::|  : | |
Db        103 -------------VIPTPGKKG--------------AAQAKALVPTPGKKGAATPAKGA 134

Qy        316 SSSLSASTSVSDSSQKKEEHNYSLFVSDN-----------LGEQPTKCSP----EEDEED 360
               :  :|    ||  : :|: : |    |:             | :| | :|     |||||
Db        135 KNGKNAKKEDSDEDEDEEDEDDSDEDEDDEEEDEFEPPIVKGVKPAKAAPAAPASEDEED 194

Qy        361 EEDVDDEDHDEGFGSEHELSENEEE----------------------EEEEEDYEDDK 396
              :|| |||: |:    | |  ::|||                      |||:| |:|:
Db        195 DEDEDDEEDDD----EEEEDDSEEEVMEITTAKGKKTPAKVVPMKAKSVAEEEDDEEEDE 250

Qy        397 DDDISDTFSEPGYENDSVEDLKEVTSISSRKRGK-----------RRYFWEYSEQLTP 443
              ||:  |   |   ::|  |: :|    :  || |           ::   | ||  ||
Db        251 DDEDEDDEEEDDEDDDEEEEEEEPVKAAPGKRKKEMTKQKEAPEAKKQKVEGSEPTTP 308


RESULT 2
JH0148
nucleolin - rat
C;Species: Rattus norvegicus (Norway rat)
C;Date: 31-Dec-1991 #sequence_revision 31-Dec-1991 #text_change 23-Jul-1999
C;Accession: JH0148; A24088; I63130
R;Bourbon, H.M.; Amalric, F.
Gene 88, 187-196, 1990
A;Title: Nucleolin gene organization in rodents: highly conserved sequences within three
A;Reference number: JH0148; MUID:90269607; PMID:2347493
A;Accession: JH0148
A;Molecule type: DNA
A;Residues: 1-712 <BOU>
A;Cross-references: UNIPARC:UPI000017723A
A;Note: the authors translated the initiation codon GTG for residue 1 as Val
R;Lischwe, M.A.; Cook, R.G.; Ahn, Y.S.; Yeoman, L.C.; Busch, H.
Biochemistry 24, 6025-6028, 1985
A;Title: Clustering of glycine and Ng,Ng-dimethylarginine in nucleolar protein C23.
A;Reference number: A24088; MUID:86104094; PMID:4084504
A;Accession: A24088
A;Molecule type: protein
A;Residues: 651-703 <LIS>
A;Cross-references: UNIPARC:UPI000017723B
R;Bourbon, H.
Gene 68, 73-84, 1988
A;Title: Sequence and structure of the nucleolin promoter in rodents: Characterization c
A;Reference number: I48118; MUID:89121496; PMID:2906027
A;Accession: I63130
A;Status: preliminary; translated from GB/EMBL/DDBJ
A;Molecule type: DNA
A;Residues: 'MV',2-44 <RES>
A;Cross-references: UNIPARC:UPI0000170A72; GB:M22090; NID:g205793; PIDN:AAA41733.1; PID:
C;Comment: This protein is the major nucleolar-specific protein in eukaryotic exponentia
C;Genetics:
A;Gene: nuc
A;Start codon: GTG
A;Introns: 5/3; 44/3; 211/1; 274/1; 303/1; 350/2; 392/1; 433/2; 485/1; 526/2; 571/1; 613
C;Superfamily: nucleolin; ribonucleoprotein repeat homology
C;Keywords: DNA binding; nucleus

F;311-376/Domain: ribonucleoprotein repeat homology <RRM1>
F;397-459/Domain: ribonucleoprotein repeat homology <RRM2>
F;489-552/Domain: ribonucleoprotein repeat homology <RRM3>
F;575-639/Domain: ribonucleoprotein repeat homology <RRM4>

  Query Match             7.8%;  Score 185.5;  DB 2;  Length 712;
  Best Local Similarity   23.3%;  Pred. No. 0.003;
  Matches   84;  Conservative   34;  Mismatches  134;  Indels  109;  Gaps   11;

Qy        129 DEEVISKTPTLAQLNSEDSQSVSDSLYYPDSLFSVKQNPLPSSFPGKKITSRAAAPVCSS 188
              : : ::  |   : :||| :   |     |     ::  :|    ||| |:  |  |  |
Db         12 ESKKMAPPPKEVEEDSEDEEMSEDE----DDSSGEEEVVIPQK-KGKKATTTPAKKVVVS 66

Qy        189 KTLQAEVPLSDCVQKASKPPSSTQIMVKTNMYHNEKVNFHVECKDYVKKAKVKINPVQQS 248
              :| : ||    :||:  |                |       |  |  |||   |
Db         67 QTKKAAVPTP--AKKAAVTPG-------------KKAAATPAKKAVTPAKVVPTP---- 106

Qy        249 RPLLSQIHTDAAKENTCYCGAVAKRQEKKGMEPLQGHATPALPFKETQELLLSPLPQEGP 308
                      |:      |:     |||       |||   |
Db        107 -----------GKKGAAQAKALVPTPGKKGA------VTPAKGAKN-------------- 135

Qy        309 GSLAAGESSSLSASTSVSDSSQKKEEHNYSLFVSDNLGEQPTKCSP------EEDEEDEE 362
              |  |  | |           | | : |:           | :| | :|      :|||||::
Db        136 GKNAKKEDSDEDEDEEDDSDEDEDEEDEFEPPVVKGVKPAKAAPAAPASEDEDEEDDD 195

Qy        363 DVDDEDHDE--------------------------------GFGSEHELSENEEEEEEE 389
              | ||:| ||                                |  |  |::|:|||:
Db        196 DEDDDDDEEEEEEDDSEEEVMEITPAKGKKTPAKVVPVKAKSVAEEEEDDEDDEDEEED 255

Qy        390 EDYEDDKDDDISDTFSE-------PGYENDSVEDLKEVTSISSRKRGKRRYFWEYSEQLT 442
              || ||::||:  |   |       ||     :  ||      :|       | ||  |
Db        256 EDEEDEEDDEDEDEEEEEEPVKAAPGKRKKEMTKQKEAPEAKKQK-------IEGSEPTT 308

Qy        443 P 443
              |
Db        309 P 309


RESULT 3
T29285
hypothetical protein C34D4.14 - Caenorhabditis elegans
C;Species: Caenorhabditis elegans
C;Date: 15-Oct-1999 #sequence_revision 15-Oct-1999 #text_change 09-Jul-2004
C;Accession: T29285
R;Du, Z.; Le, T.T.
submitted to the EMBL Data Library, May 1996
A;Description: The sequence of C. elegans cosmid C34D4.
A;Reference number: Z20600
A;Accession: T29285
A;Status: preliminary; translated from GB/EMBL/DDBJ
A;Molecule type: DNA
A;Residues: 1-2761 <DUZ>
A;Cross-references: UNIPROT:Q18447; UNIPARC:UPI000007762A; EMBL:U58755; PIDN:AAB00699.1;
A;Experimental source: strain Bristol N2; clone C34D4
C;Genetics:
A;Gene: CESP:C34D4.14
A;Map position: 4
A;Introns: 120/1; 201/3; 365/3; 445/2; 509/1; 728/3; 856/3; 902/3; 940/3; 1027/3; 1508/3;

  Query Match             7.6%;  Score 178.5;  DB 2;  Length 2761;
  Best Local Similarity   21.6%;  Pred. No. 0.039;
  Matches   84;  Conservative   53;  Mismatches  116;  Indels  135;  Gaps   15;

Qy        170 SSFPGKKI----TSRAAAPVCSSKTLQAEVPLSDCVQKASKPPSSTQIMVKTNMYHNEKV 225
              | ||   :    :|:: :|  :|: :     :|        ||||:       :      :
Db       1609 SPFPNLPVPPWRSSKSTSPAIASRLINTTTTVSGTSSPPPPPSSSTFSSLASGL-GFGL 1667

Qy        226 NFHVECKDYVKKAKVKINPVQQSRPLLSQIHTDAAKENTCYCGA---------------- 269
              | |                 | ::|  | :   |: :||   |
Db       1668 NRH---------------KQHNKPAASALSRFASVKNTTPAGTPSSGGSSGGAIGKKSM 1711

Qy        270 -----VAKRQEKKG-----------MEPLQGHATPALPFKETQELLLSPLP--------- 304
                   | :||:  |           | || | ||:|     : ||   :|
```

```
Db       1712 STTNLVDERQKTSGPSVASTGQAASAESLQ-HQTPSL-----ENLLARAMPHAFGRIAEN 1765

Qy        305 QEGPGSLAAGESSSLSAS----------TSVSDSSQKKEEHNYSLFVSDNLGEQPTKCSP 354

Db       1766 QEPEDEPMGGEESDSAASMRSAASSNSQMSMGSSSQQQQQQDSDMTPRDSAG---TPSTP 1822

Qy        355 EEDE--------------------EDEEDVDDED---HDEGFGSEHELSENEEEEEEEEE 390

Db       1823 RDDKNQTLSVSAPDLAAARQRQASAETDGDADADETNSEDKTVGADDAMEEDDEEEETME 1882

Qy        391 DYEDDKDDDISDTFSEP--------GYENDSVEDLKEV-----------------TSISS 425

Db       1883 DEEDDDDDDDDESSNENQEKLVELLGGERGLFDKLKEVITGESLSDASSSAKDATTNEAQ 1942

Qy        426 RKRGKRRYFW-----EYSEQLTPSQQER 448

Db       1943 KKGGKKPKKWFKKMSSYTDVLKGLMQNR 1970


RESULT 4
A54514
glutamic acid-rich protein precursor - malaria parasite (Plasmodium falciparum)
N;Alternate names: GARP
C;Species: Plasmodium falciparum
C;Date: 15-Oct-1994 #sequence_revision 15-Oct-1994 #text_change 09-Jul-2004
C;Accession: A54514
R;Triglia, T.; Stahl, H.D.; Crewther, P.E.; Silva, A.; Anders, R.F.; Kemp, D.J.
Mol. Biochem. Parasitol. 31, 199-202, 1988
A;Title: Structure of a Plasmodium falciparum gene that encodes a glutamic acid-rich pro
A;Reference number: A54514; MUID:89040048; PMID:2903445
A;Accession: A54514
A;Status: preliminary
A;Molecule type: DNA
A;Residues: 1-678 <TRI>
A;Cross-references: UNIPROT:P13816; UNIPARC:UPI000012B0E0; GB:J03998; NID:g160298; PIDN:
C;Genetics:
A;Introns: 25/3
C;Superfamily: histone H1
C;Keywords: tandem repeat

  Query Match             7.0%;  Score 165;  DB 2;  Length 678;
  Best Local Similarity   22.3%;  Pred. No. 0.041;
  Matches   61;  Conservative   44;  Mismatches  101;  Indels   68;  Gaps    9;

Qy        174 GKKITSRAAAPVCSSKTLQAEV-PLSDCVQKASKPPSSTQIMVKTNMYHNEKVNFHVE-- 230

Db        454 GVEIIINLEDKEACEEQHITVESRPLSQPQCKLIDEPEQLTLMDKSKV---EEKNLSIQEQ 510

Qy        231 ------------CKDYVKKAKVKINPVQQSRPLLSQIHTDAAKENTCYCGAVAKRQEKKG 278

Db        511 LIGTIGRVNVVPRRDNHKKKMAKIEEAE----LQKQKHVDKEED---------KKEESKE 557

Qy        279 MEPLQGHATPALPFKETQELLLSPLPQEGPGSLAAGESSSLSASTSVSDSSQKKEEHNYS 338

Db        558 VQE---------ESKEVQE----------------DEEEVEEDEEEEEEEEEEEEE-- 589

Qy        339 LFVSDNLGEQPTKCSPEEDEEDEEDVDDEDHDEGFGSEHELSENEEEEEEEEDYEDDKDD 398

Db        590 --------EEEEEEEEEEEEEEEEDEDEEDEDDAEEDEDDAEEDEDDAEEDDDEEDDDEE 641

Qy        399 DISDTFSEPGYENDSVEDLKEVTSISSRKRGKRR 432

Db        642 D-DDEDEDEDEEDEEEEEEEEEESEKKIKRNLRK 674


RESULT 5
A35804
nucleolin - human
N;Alternate names: phosphoprotein pp100; protein B50; protein C23
C;Species: Homo sapiens (man)
C;Date: 23-Oct-1990 #sequence_revision 23-Oct-1990 #text_change 09-Jul-2004
C;Accession: A35804; S04631; A48138; A55996
R;Srivastava, M.; McBride, O.W.; Fleming, P.J.; Pollard, H.B.; Burns, A.L.
J. Biol. Chem. 265, 14922-14931, 1990
A;Title: Genomic organization and chromosomal localization of the human nucleolin gene.
A;Reference number: A35804; MUID:90368666; PMID:2394707
A;Accession: A35804
A;Molecule type: DNA
A;Residues: 1-707 <SRI>
A;Cross-references: UNIPROT:P19338; UNIPARC:UPI0000161B7E; GB:M60858; GB:J05584; NID:g189
R;Srivastava, M.; Fleming, P.J.; Pollard, H.B.; Burns, A.L.
FEBS Lett. 250, 99-105, 1989
A;Title: Cloning and sequencing of the human nucleolin cDNA.
A;Reference number: S04631; MUID:89290043; PMID:2737305
A;Accession: S04631
A;Status: not compared with conceptual translation
A;Molecule type: mRNA
A;Residues: 1-707 <SR2>
A;Cross-references: UNIPARC:UPI0000161B7E
R;Ishikawa, F.; Matunis, M.J.; Dreyfuss, G.; Cech, T.R.
Mol. Cell. Biol. 13, 4301-4310, 1993
A;Title: Nuclear proteins that bind the pre-mRNA 3' splice site sequence r(UUAG/G) and th
A;Reference number: A48138; MUID:93309464; PMID:8321232
A;Accession: A48138
A;Molecule type: protein
A;Residues: 458-474 <ISH>
A;Cross-references: UNIPARC:UPI0000073927
A;Experimental source: HeLa cell nuclei
A;Note: sequence extracted from NCBI backbone (NCBIP:134645)
R;Jordan, P.; Heid, H.; Kinzel, V.; Kuebler, D.
Biochemistry 33, 14696-14706, 1994
A;Title: Major cell surface-located protein substrates of an ecto-protein kinase are homo
A;Reference number: A55996; MUID:95086063; PMID:7993898
A;Accession: A55996
A;Molecule type: protein
A;Residues: 231-236;349-362;399-403;458-461;655-656,'X',658-660 <JOR>
A;Cross-references: UNIPARC:UPI000017723C; UNIPARC:UPI000017723D; UNIPARC:UPI000017723E;
A;Experimental source: surface-labelled HeLa cells
C;Genetics:
A;Gene: GDB:NCL
A;Cross-references: GDB:125908; OMIM:164035
A;Map position: 2q12-2qter
C;Superfamily: nucleolin; ribonucleoprotein repeat homology
C;Keywords: DNA binding; nucleolus; nucleus; phosphoprotein; RNA binding
F;308-373/Domain: ribonucleoprotein repeat homology <RRM1>
F;394-456/Domain: ribonucleoprotein repeat homology <RRM2>
F;487-550/Domain: ribonucleoprotein repeat homology <RRM3>
F;573-634/Domain: ribonucleoprotein repeat homology <RRM4>

  Query Match             7.0%;  Score 165;  DB 2;  Length 707;
  Best Local Similarity   22.0%;  Pred. No. 0.043;
  Matches   76;  Conservative   52;  Mismatches  123;  Indels   94;  Gaps   12;

Qy        129 DEEVISKTPTLAQLNSEDSQSVSDSLYYPDSLFSVKQNPLPSSFPGKKITSRAAAPVCSS 188

Db         13 DPKKMAPPPKEVEEDSEDEEMSEDE--EDDS--SGEEVVIPQK-KGKKAAATSAKKVVVS 67

Qy        189 KTLQAEVPLSDCVQKASKPPSSTQIMVKTNMYHNEKVNFHVECKDYVKKAKVKINPVQQS 248

Db         68 PT--KKVAVATPAKKAAVTPGK--------------------KAAATPAKKTVTPAK-- 102

Qy        249 RPLLSQIHTDAAKENTCYCGAVAKRQEKKGMEPLQGHATPALPFKETQELLLSPLPQEGP 308

Db        103 ---------------------AVTTPGKKGATP--GKALVATPGKKG-----AAIPAKGA 134

Qy        309 GSLAAGESSSLSASTSVSDSSQKKEEHNYSLFVSDNLGEQP--------TKCSPEEDEED 360

Db        135 KN---GKNAKKEDSDEEEDDDSEEDEEDDEDEDEDEDEIEPAAMKAAAAAPASEDEDDED 191

Qy        361 EEDVDDEDHDEGFGSEHE----------------------LSENEEEEEEEEDYEDDKDD 398

Db        192 DEDDEDDDDDEEDDSEEEAMETTPAKGKKAAKVVPVKAKNVAEDEDEEEDDEDEDDDDE 251

Qy        399 DISDTFSEPGYENDSVED---LKEVTSISSRKRGKRRYFWEYSEQ 440

Db        252 DDEDDDDEDDEEEEEEEEEEPVKEAPGKRKKEMAKQKAAPEAKKQ 296
```

```
RESULT 6
T46637
transcription factor 1, neural - rat
N;Alternate names: neural zinc finger factor-1
C;Species: Rattus norvegicus (Norway rat)
C;Date: 18-Feb-2000 #sequence_revision 18-Feb-2000 #text_change 09-Jul-2004
C;Accession: T46637
R;Jiang, Y.; Yu, V.C.; Buchholz, F.; O'Connell, S.; Rhodes, S.J.; Candeloro, C.; Xia, Y.
J. Biol. Chem. 271, 10723-10730, 1996
A;Title: A novel family of Cys-Cys, His-Cys zinc finger transcription factors expressed
A;Reference number: A58210; MUID:96210003; PMID:8631881
A;Accession: T46637
A;Status: preliminary; translated from GB/EMBL/DDBJ
A;Molecule type: mRNA
A;Residues: 1-1187 <JIA>
A;Cross-references: UNIPROT:P70475; UNIPARC:UPI00000E86A7; EMBL:U48809; NID:g1511631; PI
C;Genetics:
A;Gene: NZF-1
C;Function:
A;Description: binds specifically to a cis-regulatory element of the beta-retinoic acid
e nervous system and in the pituitary gland
C;Keywords: DNA binding; transcription factor

  Query Match             6.9%;  Score 162;  DB 2;  Length 1187;
  Best Local Similarity   28.3%;  Pred. No. 0.12;
  Matches   52;  Conservative   35;  Mismatches   65;  Indels   32;  Gaps    7;

Qy        260 AKENTCYCGAVAKRQEKKGMEPLQGHATPALPFKETQELLLSPLPQEGPGSLAAGESSSL 319
              |:  : |   :||::: :  :|               ||    | |:  | :: | :|||:
Db         47 ARHRSVYGCPLAKKRKTQDKQP-------------QE----PAPKRKPFAVKA-DSSSV 87

Qy        320 SA--STSVSDSSQKKEEHNYSLFVSDNLGEQPTKCSPEEDEEDEEDV------DDEDHDE 371
                 :  ::    ||| :   |  ||  :       | | ||||::      |||| |:
Db         88 DECYESDGTEDMDDKEEDDDEEFSEDNDEQGDDDDEDEVDREDEEEIEEEDDEDDEDDDD 147

Qy        372 GFGSEHELSENEEEEEEEEDYEDDKDDDISDT--FSEPGYE---NDSVEDLKEVTSISSR 426
              |   | |  :::|||||||: |:::|  :| |    :|  :   ||  ::  |: : |
Db        148 GDDVEEEEDDDDEEEEEEEEEEENEDHQMSCTRIMQDPEKDDNNNDEYDNYDELVAKSLL 207

Qy        427 KRGK 430
                ||
Db        208 NLGK 211


RESULT 7
S39835
hypothetical protein YBL052c - yeast (Saccharomyces cerevisiae)
N;Alternate names: hypothetical protein YBL0507; hypothetical protein YBL0515
C;Species: Saccharomyces cerevisiae
C;Date: 16-Dec-1993 #sequence_revision 09-Sep-1994 #text_change 09-Jul-2004
C;Accession: S39835; S45787; S37336
R;Scherens, B.; el Bakkoury, M.; Vierendeels, F.; Dubois, E.; Messenguy, F.
Yeast 9, 1355-1371, 1993
A;Title: Sequencing and functional analysis of a 32 560 bp segment on the left arm of ye
A;Reference number: S39824; MUID:94205266; PMID:8154187
A;Accession: S39835
A;Molecule type: DNA
A;Residues: 1-831 <SCH>
A;Cross-references: UNIPROT:P34218; UNIPARC:UPI0000135570; EMBL:Z23261; NID:g313733; PID
A;Experimental source: strain S288C
R;Dubois, E.; El Bakkoury, M.; Glansdorff, N.; Messenguy, F.; Pierard, A.; Scherens, B.;
submitted to the Protein Sequence Database, August 1994
A;Reference number: S45782
A;Accession: S45787
A;Molecule type: DNA
A;Residues: 1-831 <DUB>
A;Cross-references: UNIPARC:UPI0000135570; EMBL:Z35814; NID:g536078; PIDN:CAA84873.1; PI
A;Experimental source: strain S288C
C;Genetics:
A;Gene: SGD:SAS3
A;Cross-references: SGD:S0000148; MIPS:YBL052c
A;Map position: 2L

  Query Match             6.8%;  Score 161;  DB 2;  Length 831;
  Best Local Similarity   20.6%;  Pred. No. 0.089;
  Matches   86;  Conservative   62;  Mismatches  116;  Indels  154;  Gaps   20;

Qy         86 SNWEQWDTY----CEDLTKYTKLTSCDIWGTKEVDYLGLDDFSSPYQDEEVISKTPTLAQ 141
              || :  ||:      |: | | :   |:         ||:      |:      || :|:
Db        486 SNNKNEDTFQQVSLNDIAKLTGMIPTDV-------VFGLEQLQVLYR-----HKTRSLSS 533

Qy        142 LN------SEDSQSVSDSLY-------YP----DSL----------FSVK--QNPLPSSF 172
              |:        || :  :::|       ||    | |          | :    |  |::
Db        534 LDDFNYIIKIDSWNRIENIYKTWSSKNYPRVKYDKLLWEPIILGPSFGINGMMNLEPTAL 593

Qy        173 PGKKITSRAAAPVCSSKT---------------------------------LQAEVPLS 198
              : :|:   ||| |: |                                  :: ||| :
Db        594 ADEALTNETMAPVISNNTHIENYNNSRAHNKRRRRRRRSSEHKTSKLHVNNIIEPEVPAT 653

Qy        199 DCVQKASKPPSSTQIMVKTNMYHNEKVNFHVECKDYVKKAKVKINPVQQSRPLLSQIHTD 258
              |  :     | |: |      :|::: :  |                  : :|  ||
Db        654 DFFEDTVS--SLTEYMCDYKNTNNDRLIYQAE-----------------KRVLESIH-- 691

Qy        259 AAKENTCYCGAVAKRQEKKGMEPLQGHATPALPFK-ETQ-ELLLSPLPQEGPGSLAAGES 316
                          ::||:        |   |  ||  ||  :    | |      |
Db        692 ---------------DRKGI--------PRSKFSTETHWELCFTIKNSETP----LGNH 723

Qy        317 SSLSASTSVSDSSQKKEEHNY--SLFVSDNLGEQPTKCSPEEDEEDEEDVDDEDHDEGFG 374
              ::    | :|    | : |::    |:| :|  |        :|||:::  :|||  ||
Db        724 AARRNDTGISSLEQDEVENDVDTELYVGENAKE-------DEDEDEDFTLDDDIEDEQI- 775

Qy        375 SEHELSENEEEEEEEEDYEDDKDDDISDTFSEPGYENDSVEDLKEVTSISSRKRGKRR 432
                      :|| :|||: ||:|  |||       | |       :|  :: |   ::| ::|
Db        776 -------SEENDEEEDTYEEDSDDD------EDGKRKGQEQDENDIESHIRKERVRKR 820


RESULT 8
T16703
hypothetical protein R07G3.3 - Caenorhabditis elegans
C;Species: Caenorhabditis elegans
C;Date: 20-Sep-1999 #sequence_revision 20-Sep-1999 #text_change 09-Jul-2004
C;Accession: T16703
R;Connell, M.
submitted to the EMBL Data Library, July 1995
A;Description: The sequence of C. elegans cosmid R07G3.
A;Reference number: Z18562
A;Accession: T16703
A;Status: preliminary; translated from GB/EMBL/DDBJ
A;Molecule type: DNA
A;Residues: 1-1974 <CON>
A;Cross-references: UNIPROT:Q09612; UNIPARC:UPI000008390B; EMBL:U23452; NID:g733564; PID:
A;Experimental source: strain Bristol N2
C;Genetics:
A;Gene: CESP:R07G3.3
A;Introns: 91/3; 158/3; 244/3; 480/3; 574/3; 722/2; 759/3; 913/2; 948/2; 1114/3; 1323/2;

  Query Match             6.8%;  Score 160.5;  DB 2;  Length 1974;
  Best Local Similarity   21.6%;  Pred. No. 0.27;
  Matches  104;  Conservative   51;  Mismatches  152;  Indels  175;  Gaps   19;

Qy         54 NPRDNFLSLEDCKDIENLE-------SFTDVLDNEGALTSNWEQWDTYCEDLTKYTKLTS 106
              | : | |  |: |  ||:        | |||    |
Db       1400 NDQINQLKQENLKLTENIRMAQLQSVSATDVESKPGP---------------------- 1436

Qy        107 CDIWGTKEVDYLGLDDFSSPYQDEEVISKTPT--LAQLNSEDSQSVSDSLYYPDSLFSVK 164
                                 ||  :    | :|||  |  |:|   |         ||    :|
Db       1437 ----------------SSASKSVSSIRQTPTKVLDPLSSAAKQPNE-----PDQTTGLK 1474

Qy        165 QNPLPSSFPGKKITSRAAAPVCSSKTLQAEVPLSDCVQKASK-----PPS---------- 209
                : | ||  |: :    |   |:    |  |     || |      |||
Db       1475 TSQQPPSFAAKRPSFGRQAQSSSNVIPPAPAPTPTSQQKVSPVKRPIPPSIPNEPLDIIP 1534

Qy        210 ---STQI---MVKTNMY-------HNEKVNFHVECKD-YVKKAKVKINPVQQSRPLLSQI 255
                  |  |      || :       |  : :  |  :  :   :   : |  : : :|  |
```

```
Db       1535 PVPSDNIPDPTPPTNSFGTVLPVPHTFQTSVRVPTQSLFSSSSTTTVQPQPEKKNVLPSI 1594

Qy        256 HTDAAKENTCYCGAVAKRQEKKGMEPLQ------GHATPALPFKETQELLLSPLPQEGPG 309
              |:|       |  :       | | |      |:    :|   |         ||:|
Db       1595 --DSAPSTP--GGNSSMVTTTSSMAPGQSIFGNIGN----VPVPTTAPTDNLALPEESVI 1646

Qy        310 SLAAGESSSLSAS-------------------------TSVSDSSQKKEEHNYSLFVSD 343
              :||:|| :| |                         | ||  :|:  ::: |  :
Db       1647 EGSAGQSSLVSGSIDQRKVQDIDLVANDGESRDSTNVGGVSSSDVRRKRTANDFELSEAK 1706

Qy        344 NLGEQPTKC------SPEEDEED-----------------------------------E 361
              | | | :        || |::|                                    |
Db       1707 RLRESPNETVTSIADIPELDDDDGVLGMEHEVSDEDPNDNTIQEQRPDVIDLENDEEVLE 1766

Qy        362 EDVDDEDHDEGFGSEHELSENEEEEEEEEDYEDDKDDDI---SDTFSEPGYENDSVEDLK 418
              :::|:|: |: ||::      || |:|||  ||| |||:   ||   ||  :||  | |
Db       1767 DEMDEEEDDDSFGND------EEFEDEEEIPEDDDDDDVVVLSDGDDEPANDNDE-ESLN 1819

Qy        419 EV 420
              ::
Db       1820 DI 1821
```

RESULT 9
G86266
hypothetical protein F3F19.25 - Arabidopsis thaliana
C;Species: Arabidopsis thaliana (mouse-ear cress)
C;Date: 02-Mar-2001 #sequence_revision 02-Mar-2001 #text_change 09-Jul-2004
C;Accession: G86266
R;Theologis, A.; Ecker, J.R.; Palm, C.J.; Federspiel, N.A.; Kaul, S.; White, O.; Alonso,
Chin, C.W.; Chung, M.K.; Conn, L.; Conway, A.B.; Conway, A.R.; Creasy, T.H.; Dewar, K.;
ansen, N.F.; Hughes, B.; Huizar, L.
Nature 408, 816-820, 2000
A;Authors: Hunter, J.L.; Jenkins, J.; Johnson-Hopson, C.; Khan, S.; Khaykin, E.; Kim, C.
C.A.; Li, J.H.; Li, Y.; Lin, X.; Liu, S.X.; Liu, Z.A.; Luros, J.S.; Maiti, R.; Marziali,
Rizzo, M.; Rooney, T.; Rowley, D.; Sakano, H.
A;Authors: Salzberg, S.L.; Schwartz, J.R.; Shinn, P.; Southwick, A.M.; Sun, H.; Tallon,
ker, M.; Wu, D.; Yu, G.; Fraser, C.M.; Venter, J.C.; Davis, R.W.
A;Title: Sequence and analysis of chromosome 1 of the plant Arabidopsis.
A;Reference number: A86141; MUID:21016719; PMID:11130712
A;Accession: G86266
A;Status: preliminary
A;Molecule type: DNA
A;Residues: 1-1128 <STO>
A;Cross-references: UNIPROT:Q9SAF6; UNIPARC:UPI00000A7DC7; GB:AE005172; NID:g4850405; PI
C;Genetics:
A;Map position: 1

Query Match 6.7%; Score 159.5; DB 2; Length 1128;
Best Local Similarity 23.3%; Pred. No. 0.16;
Matches 108; Conservative 65; Mismatches 176; Indels 115; Gaps 23;

```
Qy         33 LANSSDPDFMYE-LDREMNYQQNPRDNFLSLEDCKDIENLESFTDVL-DNEGALTSNWEQ 90
              :|  || :   : |:|:||    : |    :|:  |: | :   | | : :  : :: :
Db        678 MAELSDINHQKQALNREMEEMMSKRS---ALQ--KESEEIAKHKDKLKEQQVEMHNDISE 732

Qy         91 WDTYCEDLTK---------------YTKLTSCDIWGTKEVDYLGLDDFSSPYQDEEVISK 135
                |   :| |                    ||  |    |   |:: | |   |  || |
Db        733 LSTLSINLKKRREVFGRERSRFLAFVQKLKDCGSCGQLVNDFV-LSDLQLPSNDEVAI-- 789

Qy        136 TPTLAQLNS-EDSQSVSDSLYYPDSLFSVKQNPLPSSFPGKKITSRAAAPVCS-SK---- 189
              | :  ||    | : |||      ||          |  |  |  :  : : | ||
Db        790 LPPIGVLNDLPGSSNASDSCNIKKSLDGDASGSGGSRRPSMSILQKCTSIIFSPSKRVEH 849

Qy        190 ---TLQAEVPLSDCV-----QKASKP--------PSSTQIMVKTNMYHNEKVNFHVE--- 230
                | :  |  ||  |       |  ||         |||: |  :   |  : :|    |
Db        850 GIDTGKPEQRLSSSVAVGMETKGEKPLPVDLRLRPSSSSIPEEDEEYTDSRVQETSEGSQ 909

Qy        231 ---------CKDYVKKAKVKINPV---------QQSRPLLSQIHTDAAKENT-------- 264
                      :   :|||  :||          : |:  ||  :  :|: |
Db        910 LSEFQSSRRGRGRPRKAKPALNPTSSVKHASLEESSKDELSGHVSVTSKKTTGGGGRKRQ 969

Qy        265 ----CYCGAVAKRQEKKGMEPLQGHATPA-----LPFKETQELLLSPLPQEGPGSLAAGE 315
                  |  ::||:   : |     ||       |  |:|    :  :| :   : ||||
Db        970 HIDDTATGGKRRRQQTVAVLP----QTPGQRHYNLRRKKT----VDQVPADVEDNAAAGE 1021

Qy        316 -SSSLSASTSVSDSSQ---------KKEEHNYSLFVSDNLGEQP-TKCSPEEDEEDEEDV 364
               : ::||    |: :          :: | |  :  ::| |: |     |  :|:  ||
Db       1022 DDADIAASAPSKDTVEETVVETLRARRIETNADVVSAENNGDVPVANVEPTVNEDTNEDG 1081

Qy        365 DDEDHDEGFGSEHELSENEEEEEEEEDYEDDKDDDISDTFSEPG 408
              |:|      . | |  ::: || :::| :|| ||| |      ||
Db       1082 DEE--------EDEAQDDDNEENQDDDDDDDGDDDGS---PRPG 1114
```

RESULT 10
S38149
SIS2 protein - yeast (Saccharomyces cerevisiae)
N;Alternate names: protein YKR072c
C;Species: Saccharomyces cerevisiae
C;Date: 03-May-1994 #sequence_revision 03-May-1994 #text_change 09-Jul-2004
C;Accession: S38149; S54982; S43078
R;Pohl, T.M.; Pohl, F.M.
submitted to the Protein Sequence Database, March 1994
A;Reference number: S37897
A;Accession: S38149
A;Molecule type: DNA
A;Residues: 1-562 <POH>
A;Cross-references: UNIPROT:P36024; UNIPARC:UPI00001359B5; EMBL:Z28297; NID:g486544; PID:
A;Experimental source: strain S288C
R;di Como, C.J.; Bose, R.; Arndt, K.T.
Genetics 139, 95-107, 1995
A;Title: Overexpression of SIS2, which contains an extremely acidic region, increases the
A;Reference number: S54982; MUID:95220693; PMID:7705654
A;Accession: S54982
A;Molecule type: DNA
A;Residues: 1-562 <DIW>
A;Cross-references: UNIPARC:UPI00001359B5; EMBL:U01878; NID:g430983; PIDN:AAA80000.1; PII
C;Genetics:
A;Gene: SGD:SIS2
A;Cross-references: SGD:S0001780; MIPS:YKR072c
A;Map position: 11R
C;Keywords: nucleus

Query Match 6.6%; Score 157; DB 2; Length 562;
Best Local Similarity 20.5%; Pred. No. 0.093;
Matches 86; Conservative 64; Mismatches 120; Indels 150; Gaps 15;

```
Qy        113 KEVDYLGLDDF---SSPY--QDEEVISKTPTLAQLNSEDSQSVSDSLYYPDSLFS----- 162
              | || :  |      |:|   :: :| :| ||   |  ||:|  ::::    :: |
Db        174 KIVDTMKPDKIMATSTPISRENNKVTAKAPTSITLRKEDAQDQANNVSGQINVRSTPEET 233

Qy        163 -VKQNPLPSSFPGKKITSRAAAPVCSSKTLQAEVPLSD----CVQKASKPPSSTQI--MV 215
               |||: :||  | ::           :|| |   :|  |      :  |:   |  :|  |:
Db        234 PVKQSVIPSIIPKRE----------NSKNLDPRLPQDDGKLHVLFGATGSLSVFKIKPMI 283

Qy        216 K--TNMYHNEKVNFHVECKDYVKKAKVKINPVQQSRPLLSQIHTDAAKENTCYCGAVAKR 273
              |    :|  :::: |                     :|:|  |  ::   |  : |
Db        284 KKLEEIYGRDRISIQV--------------------ILTQSATQFFEQR--YTKKIIKS 320

Qy        274 QEKKGMEPLQGHATPALPFKET------------------------------------ 295
               ||   :  |  :||| |   |
Db        321 SEKLN-KMSQYESTPATPVTPTPGQCNMAQVVELPPHIQLWTDQDEWDAWKQRTDPVLHI 379

Qy        296 ------QELLLSPLPQEGPGSLAAG-------------------ESSSLSASTSVSDSS 329
                   |:::||      :| |                      : |: :||  | :
Db        380 ELRRWADILVVAPLTANTLSKIALGLCDNLLTSVIRAWNPSYPILLAPSMVSSTFNSMMT 439

Qy        330 QK-----KEEHNY------SLFVSD---------------------NLGEQPTKCSPEE 356
              :|      ||| ::      |  | |                    ||  |     |:
Db        440 KKQLQTIKEEMSWVTVFKPSEKVMDINGDIGLGGMMDWNEIVNKIVMKLGGYPKNNEEED 499

Qy        357 DEEDEEDVDDEDHDEGFGSEHELSENEEEEEEEEDYEDDKDDDISDTFSEPGYENDSVED 416
              |:||||: |||: |    :|:   ::::::::::| :|| |||  |   |   |   : |
Db        500 DDEDEEEDDDEEEDTEDKNENNNDDDDDDDDDDDDDDDDDDDDDDEDEDEAETPGIID 559
```

```
RESULT 11
T42963
hypothetical protein 48 - ateline herpesvirus 3 (strain 73)
C;Species: ateline herpesvirus 3
A;Variety: strain 73
C;Date: 21-Jan-2000 #sequence_revision 21-Jan-2000 #text_change 09-Jul-2004
C;Accession: T42963
R;Albrecht, J.C.; Fleckenstein, B.
submitted to the EMBL Data Library, August 1998
A;Description: Primary structure of the herpesvirus ateles genome.
A;Reference number: Z22274
A;Accession: T42963
A;Status: preliminary; translated from GB/EMBL/DDBJ
A;Molecule type: DNA
A;Residues: 1-792 <ALB>
A;Cross-references: UNIPROT:Q9YTL7; UNIPARC:UPI00000EC1E3; EMBL:AF083424; PIDN:AAC95573.
A;Experimental source: strain 73

  Query Match             6.5%;  Score 153;  DB 2;  Length 792;
  Best Local Similarity   22.6%;  Pred. No. 0.24;
  Matches   85;  Conservative   62;  Mismatches  153;  Indels   76;  Gaps   14;

Qy        110 WGTKEVDYLGLD--DFSSPYQDEEVISKTPTLAQLNSEDSQSV-SDSLYYPDSLFSVKQN 166
              |   |: :|  |  |||   :|  :|:       |      | : || :   ::| |   :
Db        174 WMVPEIMFLPSDYHDFSEHKKDISLIATCLNCCWLYFMTYQYMSSDLIAIEEALQSSYLS 233

Qy        167 PLPSSFPG-KKITSRAAAPVCSSKTLQAEVPLSDCV----------------------- 201
              |::|   ||     | : | : :: :| ::| :
Db        234 ICGSTYPSYSKILELLTANM-SKEHIRQKVNVTDFIKPSLHQMIRDTKKEPRQKTKTLMI 292

Qy        202 --------------QKASKPPSST---QIMVKTNMYHNEKVNFHV-----ECKDYVKKA 238
                           |   | |||      : : : :: ::| | :     : :|
Db        293 SILGSRGIGLDLFRTQDVLKFPSSDAKFMAVSQPDNFNEKEVEFSMTGGKTDSEDVTAPR 352

Qy        239 KVKINPVQQSRPLLSQIHTDAAKENTCYCGAVAK--------------RQEKKGMEPLQG 284
              ||  | :  :|  |  :  |  ::|  | |   |               : | :| :  :
Db        353 KVGKNSL--NRKYLENL-KDNKRKNNNYSGRNNKYKGDGANDKDKSIDKNESEGGDHSEI 409

Qy        285 HATPALPFKETQELLLSPLPQEGPGSLAAGESSSLSASTSVSDSSQKKEEHNYSLFVSDN 344
              :       |:      :          |: :||    |   |  ::  | || |
Db        410 NREKNRKRKKPNGFRVGDKEVGEEKSVKSGEGKK-SEKDSEEEAEDKDEEEN----KKKG 464

Qy        345 LGEQPTKCSPEEDEEDEEDVDDEDHDEGFGSEHELSENEEEEEEEEDYEDDKD-DDISDT 403
              ||:  :   |||||:||: ::|: ||    | |  |:|||||:||| ||::| :|  |
Db        465 DGEEDEEDEEEEDEEEEEE-EEEEEDEEEEEEEEDEEDEEEEEDEEDEEDEEDEEDE 523

Qy        404 FSEPGYENDSVEDLKE 419
                |   |::  |: :|
Db        524 EDEEDEEDEEEEEDEE 539


RESULT 12
S55785
nucleolar protein gar2 - fission yeast (Schizosaccharomyces pombe)
N;Alternate names: probable RNA-binding protein RBD18
C;Species: Schizosaccharomyces pombe
C;Date: 28-Oct-1996 #sequence_revision 07-Feb-1997 #text_change 05-Oct-2004
C;Accession: S55785; S68087; T37634; S52318
R;Gulli, M.P.; Girard, J.P.; Zabetakis, D.; Lapeyre, B.; Melese, T.; Caizergues-Ferrer,
Nucleic Acids Res. 23, 1912-1918, 1995
A;Title: gar2 is a nucleolar protein from Schizosaccharomyces pombe required for 18S rRN
A;Reference number: S55785; MUID:95319932; PMID:7596817
A;Accession: S55785
A;Status: nucleic acid sequence not shown
A;Molecule type: DNA
A;Residues: 1-500 <GUL>
A;Cross-references: UNIPROT:P41891; UNIPARC:UPI000016914D; EMBL:Z48166; NID:g663261; PID
R;VanHoy, R.W.; Wise, J.A.
Curr. Genet. 29, 307-315, 1996
A;Title: Molecular analysis of a novel Schizosaccharomyces pombe gene containing two RNP
A;Reference number: S68083; MUID:96171513; PMID:8598051
A;Accession: S68087
A;Status: nucleic acid sequence not shown; not compared with conceptual translation
A;Molecule type: mRNA
A;Residues: 374-407 <VAN>
A;Cross-references: UNIPARC:UPI000017A376
R;Brown, D.; Churcher, C.M.; Barrell, B.G.; Rajandream, M.A.; Wood, V.
submitted to the EMBL Data Library, September 1997
A;Reference number: Z21733
A;Accession: T37634
A;Status: preliminary; translated from GB/EMBL/DDBJ
A;Molecule type: DNA
A;Residues: 1-338,'S',340-500 <BRO>
A;Cross-references: UNIPARC:UPI000012B0DD; EMBL:Z99091; PIDN:CAB11772.1; GSPDB:GN00066; S
A;Experimental source: strain 972h-; cosmid c13F5
C;Genetics:
A;Gene: gar2; SPDB:SPAC13F5.09
A;Map position: 1
F;264-331/Domain: ribonucleoprotein repeat homology <RRM1>
F;367-433/Domain: ribonucleoprotein repeat homology <RRM2>

  Query Match             6.5%;  Score 152.5;  DB 2;  Length 500;
  Best Local Similarity   23.4%;  Pred. No. 0.15;
  Matches   75;  Conservative   38;  Mismatches  115;  Indels   93;  Gaps   10;

Qy        145 EDSQSVSDSLYYPDSLFSVKQNPLPSSFPGKKITSRAAAPVC--SSKTLQAEVPLSDCVQ 202
              :|  ||  |:    |    |:  :      ||||  ||  :  ||||        |
Db          4 KDKTSVKKSVKETAS----KKGAIEKPSKSKKITKEAAKEIAKQSSKT--------DVSP 51

Qy        203 KASKPPSSTQIMVKTNMYHNEKVNFHVECKDYVKKAKVKINPVQQSRPLLSQIHTDAAKE 262
              | ||                          |:||   :|    : :  |
Db         52 KKSK----------------------------KEAKRASSPEPSKKSVKKQ-------- 74

Qy        263 NTCYCGAVAKRQEKKGMEPLQGHATPALPFKETQELLLSPLPQEGPGSLAAGESSSLSAS 322
                        |: :||     :  :  :    |:          |   | :  ||||  :|
Db         75 ---------KKSKKKEESSSESESESSSSESESSS-------SESESSSSESESSSSESS 118

Qy        323 TSVSD------SSQKKEEHNYSLFVSDNLGEQPTKCSPEEDEEDEEDVDDEDHDEGFGSE 376
              :| |:      : :|||  : |   |::  |:       || :|   |  |        ||
Db        119 SSESEEEVIVKTEEKKESSSESSSSSESEEEEEAVVKIEEKKESSSDSSSESSSSESESE 178

Qy        377 HELSENEEEEEEEEDYEDDKD---------DDISDTFSEPGYENDSVEDLKEVTSISSRK 427
                ||:|||||  |  |: |:         :  ||: || | ::||  | :  :|    |
Db        179 SSSSESEEEEEVVEKTEEKKEGSSESSSDSESSSDSSSESG-DSDSSSDSESESSSEDEK 237

Qy        428 RGKRRYFWEYSEQLTPSQQER 448
              : |             |: :||
Db        238 KRKAE----------PASEER 248


RESULT 13
T24169
hypothetical protein R11A5.1b - Caenorhabditis elegans
C;Species: Caenorhabditis elegans
C;Date: 15-Oct-1999 #sequence_revision 15-Oct-1999 #text_change 09-Jul-2004
C;Accession: T24169
R;McMurray, A.
submitted to the EMBL Data Library, November 1996
A;Reference number: Z19848
A;Accession: T24169
A;Status: preliminary; translated from GB/EMBL/DDBJ
A;Molecule type: DNA
A;Residues: 1-896 <WIL>
A;Cross-references: UNIPROT:O62348; UNIPARC:UPI000007DBCA; EMBL:Z83122; PIDN:CAB05601.1;
A;Experimental source: clone R11A5
C;Genetics:
A;Gene: CESP:R11A5.1b
A;Map position: 1
A;Introns: 14/2; 40/3; 277/3; 308/3; 353/3; 550/1; 658/1; 798/3

  Query Match             6.5%;  Score 152.5;  DB 2;  Length 896;
  Best Local Similarity   32.3%;  Pred. No. 0.3;
  Matches   43;  Conservative   27;  Mismatches   50;  Indels   13;  Gaps    4;

Qy        296 QELLLSPLPQEGPGSLAAGESSSLSASTSVSDSSQKKEEHNYSLFVSDNLGEQPTKCSPE 355
```

```
           :|: :|  |        |    ||    || |||  :|| |  :    |:  |:  :   :
Db        531 EEIFMSKKP-------APELISSFKEPTS-SDSLLRKEAHAPNELKLDDSDEE--EDDDD 580

Qy        356 EDEEDEEDVDDEDHDEGFGSEHELSENEEEEEEEEDYEDDKDDDISDTFSEPGYENDSVE 415
              ||||::|| ::|| ||    : |  ||| |||||:| ::: :::  :   |   |:||
Db        581 EDEEEDEDEEEEDEDE---EDEEEEENESEEEEEDDEDEESEEESDEEDEEEEEEDDSEP 637

Qy        416 DLKEVTSISSRKR 428
              : |     |::|:
Db        638 EEKPNPKPSTKKQ 650


RESULT 14
T24166
hypothetical protein R11A5.1a - Caenorhabditis elegans
C;Species: Caenorhabditis elegans
C;Date: 15-Oct-1999 #sequence_revision 15-Oct-1999 #text_change 09-Jul-2004
C;Accession: T24166
R;McMurray, A.
submitted to the EMBL Data Library, November 1996
A;Reference number: Z19848
A;Accession: T24166
A;Status: preliminary; translated from GB/EMBL/DDBJ
A;Molecule type: DNA
A;Residues: 1-906 <WIL>
A;Cross-references: UNIPROT:O45718; UNIPARC:UPI0000083395; EMBL:Z83122; PIDN:CAB05598.1;
A;Experimental source: clone R11A5
C;Genetics:
A;Gene: CESP:R11A5.1a
A;Map position: 1
A;Introns: 14/2; 40/3; 277/3; 318/3; 363/3; 560/1; 668/1; 808/3

  Query Match             6.5%;  Score 152.5;  DB 2;  Length 906;
  Best Local Similarity   32.3%;  Pred. No. 0.3;
  Matches   43;  Conservative   27;  Mismatches   50;  Indels   13;  Gaps    4;

Qy        296 QELLLSPLPQEGPGSLAAGESSSLSASTSVSDSSQKKEEHNYSLFVSDNLGEQPTKCSPE 355
              :|: :|  |        |    ||    || |||  :|| |  :    |:  |:  :   :
Db        541 EEIFMSKKP-------APELISSFKEPTS-SDSLLRKEAHAPNELKLDDSDEE--EDDDD 590

Qy        356 EDEEDEEDVDDEDHDEGFGSEHELSENEEEEEEEEDYEDDKDDDISDTFSEPGYENDSVE 415
              ||||::|| ::|| ||    : |  ||| |||||:| ::: :::  :   |   |:||
Db        591 EDEEEDEDEEEEDEDE---EDEEEEENESEEEEEDDEDEESEEESDEEDEEEEEEDDSEP 647

Qy        416 DLKEVTSISSRKR 428
              : |     |::|:
Db        648 EEKPNPKPSTKKQ 660


RESULT 15
S52522
hypothetical protein YPL009c - yeast (Saccharomyces cerevisiae)
N;Alternate names: hypothetical protein LPA8c; hypothetical protein YP8132.04c
C;Species: Saccharomyces cerevisiae
C;Date: 08-May-1995 #sequence_revision 21-Jul-1995 #text_change 09-Jul-2004
C;Accession: S52522; S59684
R;Badcock, K.; Churcher, C.
submitted to the EMBL Data Library, February 1995
A;Reference number: S52519
A;Accession: S52522
A;Molecule type: DNA
A;Residues: 1-1038 <BAD>
A;Cross-references: UNIPROT:Q12532; UNIPARC:UPI0000052E40; EMBL:Z48483; NID:g683777; PID
A;Experimental source: strain AB972
R;Hall, J.; Ahmed, A.; Bussey, H.; Fortin, N.; Friesen, J.D.; Storms, R.K.; Vo, D.H.; Wa
submitted to the EMBL Data Library, August 1995
A;Description: The sequence of Saccharomyces cerevisiae chromosome XVI left arm.
A;Reference number: S59677
A;Accession: S59684
A;Molecule type: DNA
A;Residues: 1-1038 <HAL>
A;Cross-references: UNIPARC:UPI0000052E40; EMBL:U33335; NID:g965076; PID:g965084; MIPS:Y
C;Genetics:
A;Cross-references: SGD:S0005930
A;Map position: 16L

  Query Match             6.4%;  Score 152;  DB 2;  Length 1038;
  Best Local Similarity   24.8%;  Pred. No. 0.38;
  Matches   61;  Conservative   40;  Mismatches   89;  Indels   56;  Gaps   10;

Qy        201 VQKASKPPSSTQIMVK----TNMYHNEKVNFHVECKDYVKKAKVKINPVQQSRPLL---S 253
              |      |  : ||  |      ::| :   | ||  |:  :| :|  | : |:  |     |
Db        588 VMMGKSPAETDQIYSKYIEDDDIYMSNSFNSHVWIKN-PEKTEVPPNTLMQAGILCMSSS 646

Qy        254 QIHTDAAKENTCYCGAVAKRQEKKGMEPLQGHATPALP---FKETQELLLSPLPQEGPGS 310
              :  :     :  :| |        | :     |      ||   |:   |   : ||   |
Db        647 EAWSKKISSSPWWCFA-------KNVSKFDGSDNSILPEGAFRLKNENDQNHLP---PAQ 696

Qy        311 LAAGESSSLSASTSVSDSSQKKEEHNYSLFVSDNLGEQPTKCSPEEDEEDEEDVDDEDHD 370
              |  |       || :: :   :|                    ||:||:||: ::|: :
Db        697 LVMGFGFLWKVKTSGNEDNGDDDE------------------EEEEEEEEEEEEEEE 736

Qy        371 EGFGSEHELSENEEEEEEEDYEDDKDD----------DISDTFSEPGYENDSVE-DLKE 419
              |    | |  | ||||:|||  :|: |           | :|:     :|:|::| |:::
Db        737 E----EEEEEEKEEEEKEEEQQQDEDDSNEVNGLEKGGDSNDSTKNNSFEHDNLEKDIEK 792

Qy        420 VTSISS 425
                :|||
Db        793 HCTISS 798


Search completed: February 28, 2006, 08:54:18
Job time : 28.3216 secs
```

This Page Blank (uspto)

GenCore version 5.1.7

OM protein - protein search, using sw model

Run on: February 28, 2006, 08:45:31 ; Search time 164.304 Seconds
(without alignments)
1923.730 Million cell updates/sec

Title: US-10-717-665A-44_COPY_1_448
Perfect score: 2364
Sequence: 1 MPQPSVSGMDPPFGDAFRSH.....GKRRYFWEYSEQLTPSQQER 448

Scoring table: BLOSUM62
Gapop 10.0 , Gapext 0.5

Searched: 2166443 seqs, 705528306 residues

Total number of hits satisfying chosen parameters: 2166443

Minimum DB seq length: 0
Maximum DB seq length: 2000000000

Post-processing: Minimum Match 0%
Maximum Match 100%
Listing first 45 summaries

Database : UniProt_05.80:*
1: uniprot_sprot:*
2: uniprot_trembl:*

Pred. No. is the number of results predicted by chance to have a score greater than or equal to the score of the result being printed, and is derived by analysis of the total score distribution.

SUMMARIES

Result No.	Score	% Query Match	Length	DB	ID	Description
1	2356	99.7	639	2	Q8IZG1_HUMAN	Q8izg1 homo sapien
2	2356	99.7	639	2	Q8IUR6_HUMAN	Q8iur6 homo sapien
3	2353	99.5	604	2	Q5HYK0_HUMAN	Q5hyk0 homo sapien
4	2195.5	92.9	640	2	Q8CDG5_MOUSE	Q8cdg5 mus musculu
5	2148	90.9	408	2	Q5HYG4_HUMAN	Q5hyg4 homo sapien
6	2148	90.9	417	2	Q86YR3_HUMAN	Q86yr3 homo sapien
7	1704.5	72.1	351	2	Q8C8N7_MOUSE	Q8c8n7 mus musculu
8	1602.5	67.8	330	2	Q9CTQ7_MOUSE	Q9ctq7 mus musculu
9	656	27.7	131	2	Q5ZMD5_CHICK	Q5zmd5 gallus gall
10	644	27.2	600	2	Q4RRX3_TETNG	Q4rrx3 tetraodon n
11	194	8.2	976	2	Q9DUN0_HHV8	Q9dun0 human herpe
12	191.5	8.1	1003	2	Q91LX9_HHV8	Q91lx9 human herpe
13	191.5	8.1	1129	2	Q9QR71_HHV8	Q9qr71 human herpe
14	190.5	8.1	2649	2	Q7RAS7_PLAYO	Q7ras7 plasmodium
15	190	8.0	1162	2	Q98148_HHV8	Q98148 human herpe
16	189.5	8.0	707	2	Q99K50_MOUSE	Q99k50 mus musculu
17	189.5	8.0	707	2	Q8CE30_MOUSE	Q8ce30 mus musculu
18	188.5	8.0	444	2	Q9CT46_MOUSE	Q9ct46 mus musculu
19	188.5	8.0	706	1	NUCL_MOUSE	P09405 mus musculu
20	188.5	8.0	707	2	Q8CD23_MOUSE	Q8cd23 mus musculu
21	188.5	8.0	707	2	Q548M9_MOUSE	Q548m9 mus musculu
22	187.5	7.9	1089	2	O40947_HHV8	O40947 human herpe
23	185.5	7.8	712	1	NUCL_RAT	P13383 rattus norv
24	185	7.8	1036	2	Q9DUM3_HHV8	Q9dum3 human herpe
25	184	7.8	714	2	Q5U328_RAT	Q5u328 rattus norv
26	182	7.7	771	2	Q54IN8_DICDI	Q54in8 dictyosteli
27	179	7.6	416	2	Q9YPA9_HHV8	Q9ypa9 human herpe
28	179	7.6	842	2	Q76MT4_RATRT	Q76mt4 rattus ratt
29	179	7.6	1487	2	Q7SDF8_NEUCR	Q7sdf8 neurospora
30	178.5	7.6	2761	2	Q18447_CAEEL	Q18447 caenorhabdi
31	178	7.5	1018	2	Q55YM7_CRYNE	Q55ym7 cryptococcu
32	178	7.5	1018	2	Q5KMZ9_CRYNE	Q5kmz9 cryptococcu
33	178	7.5	1308	2	Q5CVU2_CRYPV	Q5cvu2 cryptospori
34	177	7.5	2750	2	Q54V59_DICDI	Q54v59 dictyosteli
35	176.5	7.5	831	2	Q7RIE7_PLAYO	Q7rie7 plasmodium
36	176.5	7.5	855	2	Q55SC9_CRYNE	Q55sc9 cryptococcu
37	176.5	7.5	1852	2	Q9C2H4_NEUCR	Q9c2h4 neurospora
38	176	7.4	391	2	Q7RPC8_PLAYO	Q7rpc8 plasmodium
39	176	7.4	2191	2	Q7RY51_NEUCR	Q7ry51 neurospora
40	173.5	7.3	676	2	Q8I510_PLAF7	Q8i510 plasmodium
41	173	7.3	793	2	Q8IBJ8_PLAF7	Q8ibj8 plasmodium
42	173	7.3	4037	2	Q4WPX3_ASPFU	Q4wpx3 aspergillus
43	172.5	7.3	755	2	Q4XSA4_PLACH	Q4xsa4 plasmodium
44	172	7.3	679	2	Q95XW8_CAEEL	Q95xw8 caenorhabdi
45	172	7.3	1156	2	Q8F2J8_LEPIN	Q8f2j8 leptospira

ALIGNMENTS

RESULT 1
Q8IZG1_HUMAN
ID Q8IZG1_HUMAN PRELIMINARY; PRT; 639 AA.
AC Q8IZG1;
DT 01-MAR-2003 (TrEMBLrel. 23, Created)
DT 01-MAR-2003 (TrEMBLrel. 23, Last sequence update)
DT 01-MAR-2004 (TrEMBLrel. 26, Last annotation update)
DE Adult retina protein.
OS Homo sapiens (Human).
OC Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;
OC Mammalia; Eutheria; Euarchontoglires; Primates; Catarrhini; Hominidae;
OC Homo.
OX NCBI_TaxID=9606;
RN [1]
RP NUCLEOTIDE SEQUENCE.
RA Li F., Yao K.T.;
RL Submitted (AUG-2002) to the EMBL/GenBank/DDBJ databases.
DR EMBL; AY139008; AAN28956.1; -; mRNA.
DR InterPro; IPR004827; TF_bZIP.
DR PROSITE; PS00036; BZIP_BASIC; UNKNOWN_1.
SQ SEQUENCE 639 AA; 72090 MW; 3A19E0926B9A6406 CRC64;

Query Match 99.7%; Score 2356; DB 2; Length 639;
Best Local Similarity 99.8%; Pred. No. 4e-113;
Matches 447; Conservative 0; Mismatches 1; Indels 0; Gaps 0;

```
Qy          1 MPQPSVSGMDPPFGDAFRSHTFSEQTLMSTDLLANSSDPDFMYELDREMNYQQNPRDNFL 60
              ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db          1 MPQPSVSGMDPPFGDAFRSHTFSEQTLMSTDLLANSSDPDFMYELDREMNYQQNPRDNFL 60

Qy         61 SLEDCKDIENLESFTDVLDNEGALTSNWEQWDTYCEDLTKYTKLTSCDIWGTKEVDYLGL 120
              ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db         61 SLEDCKDIENLESFTDVLDNEGALTSNWEQWDTYCEDLTKYTKLTSCDIWGTKEVDYLGL 120

Qy        121 DDFSSPYQDEEVISKTPTLAQLNSEDSQSVSDSLYYPDSLFSVKQNPLPSSFPGKKITSR 180
              ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db        121 DDFSSPYQDEEVISKTPTLAQLNSEDSQSVSDSLYYPDSLFSVKQNPLPSSFPGKKITSR 180

Qy        181 AAAPVCSSKTLQAEVPLSDCVQKASKPPSSTQIMVKTNMYHNEKVNFHVECKDYVKKAKV 240
              ||||||||||||||||||||||||||| ||||||||||||||||||||||||||||||||||
Db        181 AAAPVCSSKTLQAEVPLSDCVQKASKPTSSTQIMVKTNMYHNEKVNFHVECKDYVKKAKV 240

Qy        241 KINPVQQSRPLLSQIHTDAAKENTCYCGAVAKRQEKKGMEPLQGHATPALPFKETQELLL 300
              ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db        241 KINPVQQSRPLLSQIHTDAAKENTCYCGAVAKRQEKKGMEPLQGHATPALPFKETQELLL 300

Qy        301 SPLPQEGPGSLAAGESSSLSASTSVSDSSQKKEEHNYSLFVSDNLGEQPTKCSPEEDEED 360
              ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db        301 SPLPQEGPGSLAAGESSSLSASTSVSDSSQKKEEHNYSLFVSDNLGEQPTKCSPEEDEED 360

Qy        361 EEDVDDEDHDEGFGSEHELSENEEEEEEEEDYEDDKDDDISDTFSEPGYENDSVEDLKEV 420
              ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db        361 EEDVDDEDHDEGFGSEHELSENEEEEEEEEDYEDDKDDDISDTFSEPGYENDSVEDLKEV 420
```

```
Qy        421 TSISSRKRGKRRYFWEYSEQLTPSQQER 448
              ||||||||||||||||||||||||||||||
Db        421 TSISSRKRGKRRYFWEYSEQLTPSQQER 448


RESULT 2
Q8IUR6_HUMAN
ID   Q8IUR6_HUMAN PRELIMINARY;      PRT;   639 AA.
AC   Q8IUR6;
DT   01-MAR-2003 (TrEMBLrel. 23, Created)
DT   01-MAR-2003 (TrEMBLrel. 23, Last sequence update)
DT   01-MAR-2004 (TrEMBLrel. 26, Last annotation update)
DE   Adult retina protein.
GN   Name=LOC153222;
OS   Homo sapiens (Human).
OC   Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;
OC   Mammalia; Eutheria; Euarchontoglires; Primates; Catarrhini; Hominidae;
OC   Homo.
OX   NCBI_TaxID=9606;
RN   [1]
RP   NUCLEOTIDE SEQUENCE.
RC   TISSUE=Lung;
RX   MEDLINE=22388257; PubMed=12477932; DOI=10.1073/pnas.242603899;
RA   Strausberg R.L., Feingold E.A., Grouse L.H., Derge J.G.,
RA   Klausner R.D., Collins F.S., Wagner L., Shenmen C.M., Schuler G.D.,
RA   Altschul S.F., Zeeberg B., Buetow K.H., Schaefer C.F., Bhat N.K.,
RA   Hopkins R.F., Jordan H., Moore T., Max S.I., Wang J., Hsieh F.,
RA   Diatchenko L., Marusina K., Farmer A.A., Rubin G.M., Hong L.,
RA   Stapleton M., Soares M.B., Bonaldo M.F., Casavant T.L., Scheetz T.E.,
RA   Brownstein M.J., Usdin T.B., Toshiyuki S., Carninci P., Prange C.,
RA   Raha S.S., Loquellano N.A., Peters G.J., Abramson R.D., Mullahy S.J.,
RA   Bosak S.A., McEwan P.J., McKernan K.J., Malek J.A., Gunaratne P.H.,
RA   Richards S., Worley K.C., Hale S., Garcia A.M., Gay L.J., Hulyk S.W.,
RA   Villalon D.K., Muzny D.M., Sodergren E.J., Lu X., Gibbs R.A.,
RA   Fahey J., Helton E., Ketteman M., Madan A., Rodrigues S., Sanchez A.,
RA   Whiting M., Madan A., Young A.C., Shevchenko Y., Bouffard G.G.,
RA   Blakesley R.W., Touchman J.W., Green E.D., Dickson M.C.,
RA   Rodriguez A.C., Grimwood J., Schmutz J., Myers R.M.,
RA   Butterfield Y.S.N., Krzywinski M.I., Skalska U., Smailus D.E.,
RA   Schnerch A., Schein J.E., Jones S.J.M., Marra M.A.;
RT   "Generation and initial analysis of more than 15,000 full-length human
RT   and mouse cDNA sequences.";
RL   Proc. Natl. Acad. Sci. U.S.A. 99:16899-16903(2002).
RN   [2]
RP   NUCLEOTIDE SEQUENCE.
RC   TISSUE=Lung;
RA   Director MGC Project;
RL   Submitted (DEC-2002) to the EMBL/GenBank/DDBJ databases.
DR   EMBL; BC041709; AAH41709.1; -; mRNA.
DR   Ensembl; ENSG00000164463; Homo sapiens.
DR   InterPro; IPR004827; TF_bZIP.
DR   PROSITE; PS00036; BZIP_BASIC; UNKNOWN_1.
SQ   SEQUENCE   639 AA;  72118 MW;  ECFB92D9290DEDEB CRC64;

  Query Match             99.7%;  Score 2356;  DB 2;  Length 639;
  Best Local Similarity   99.8%;  Pred. No. 4e-113;
  Matches  447;  Conservative    0;  Mismatches    1;  Indels    0;  Gaps    0;

Qy          1 MPQPSVSGMDPPFGDAFRSHTFSEQTLMSTDLLANSSDPDFMYELDREMNYQQNPRDNFL 60
              ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db          1 MPQPSVSGMDPPFGDAFRSHTFSEQTLMSTDLLANSSDPDFMYELDREMNYQQNPRDNFL 60

Qy         61 SLEDCKDIENLESFTDVLDNEGALTSNWEQWDTYCEDLTKYTKLTSCDIWGTKEVDYLGL 120
              ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db         61 SLEDCKDIENLESFTDVLDNEGALTSNWEQWDTYCEDLTKYTKLTSCDIWGTKEVDYLGL 120

Qy        121 DDFSSPYQDEEVISKTPTLAQLNSEDSQSVSDSLYYPDSLFSVKQNPLPSSFPGKKITSR 180
              ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db        121 DDFSSPYQDEEVISKTPTLAQLNSEDSQSVSDSLYYPDSLFSVKQNPLPSSFPGKKITSR 180

Qy        181 AAAPVCSSKTLQAEVPLSDCVQKASKPPSSTQIMVKTNMYHNEKVNFHVECKDYVKKAKV 240
              |||||||||||||||||||||||| ||||||||||||||||||||||||||||||||||||
Db        181 AAAPVCSSKTLQAEVPLSDCVQKASKPTSSTQIMVKTNMYHNEKVNFHVECKDYVKKAKV 240

Qy        241 KINPVQQSRPLLSQIHTDAAKENTCYCGAVAKRQEKKGMEPLQGHATPALPFKETQELLL 300
              ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db        241 KINPVQQSRPLLSQIHTDAAKENTCYCGAVAKRQEKKGMEPLQGHATPALPFKETQELLL 300

Qy        301 SPLPQEGPGSLAAGESSSLSASTSVSDSSQKKEEHNYSLFVSDNLGEQPTKCSPEEDEED 360
              ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db        301 SPLPQEGPGSLAAGESSSLSASTSVSDSSQKKEEHNYSLFVSDNLGEQPTKCSPEEDEED 360

Qy        361 EEDVDDEDHDEGFGSEHELSENEEEEEEEEDYEDDKDDDISDTFSEPGYENDSVEDLKEV 420
              ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db        361 EEDVDDEDHDEGFGSEHELSENEEEEEEEEDYEDDKDDDISDTFSEPGYENDSVEDLKEV 420

Qy        421 TSISSRKRGKRRYFWEYSEQLTPSQQER 448
              ||||||||||||||||||||||||||||
Db        421 TSISSRKRGKRRYFWEYSEQLTPSQQER 448


RESULT 3
Q5HYK0_HUMAN
ID   Q5HYK0_HUMAN PRELIMINARY;      PRT;   604 AA.
AC   Q5HYK0;
DT   10-MAY-2005 (TrEMBLrel. 30, Created)
DT   10-MAY-2005 (TrEMBLrel. 30, Last sequence update)
DT   10-MAY-2005 (TrEMBLrel. 30, Last annotation update)
DE   Hypothetical protein DKFZp313F2319 (Fragment).
GN   Name=DKFZp313F2319;
OS   Homo sapiens (Human).
OC   Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;
OC   Mammalia; Eutheria; Euarchontoglires; Primates; Catarrhini; Hominidae;
OC   Homo.
OX   NCBI_TaxID=9606;
RN   [1]
RP   NUCLEOTIDE SEQUENCE.
RC   TISSUE=Adipose;
RG   The German cDNA Consortium;
RA   Koehrer K., Beyer A., Mewes H.W., Weil B., Amid C., Osanger A.,
RA   Fobo G., Han M., Wiemann S.;
RL   Submitted (JAN-2005) to the EMBL/GenBank/DDBJ databases.
DR   EMBL; BX647573; CAI46104.1; -; mRNA.
DR   InterPro; IPR004827; TF_bZIP.
DR   PROSITE; PS00036; BZIP_BASIC; UNKNOWN_1.
KW   Hypothetical protein.
FT   NON_TER     604    604
SQ   SEQUENCE   604 AA;  68572 MW;  977C229B63E2E4C2 CRC64;

  Query Match             99.5%;  Score 2353;  DB 2;  Length 604;
  Best Local Similarity   99.6%;  Pred. No. 5.4e-113;
  Matches  446;  Conservative    1;  Mismatches    1;  Indels    0;  Gaps    0;

Qy          1 MPQPSVSGMDPPFGDAFRSHTFSEQTLMSTDLLANSSDPDFMYELDREMNYQQNPRDNFL 60
              ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db          1 MPQPSVSGMDPPFGDAFRSHTFSEQTLMSTDLLANSSDPDFMYELDREMNYQQNPRDNFL 60

Qy         61 SLEDCKDIENLESFTDVLDNEGALTSNWEQWDTYCEDLTKYTKLTSCDIWGTKEVDYLGL 120
              ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db         61 SLEDCKDIENLESFTDVLDNEGALTSNWEQWDTYCEDLTKYTKLTSCDIWGTKEVDYLGL 120

Qy        121 DDFSSPYQDEEVISKTPTLAQLNSEDSQSVSDSLYYPDSLFSVKQNPLPSSFPGKKITSR 180
              ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db        121 DDFSSPYQDEEVISKTPTLAQLNSEDSQSVSDSLYYPDSLFSVKQNPLPSSFPGKKITSR 180

Qy        181 AAAPVCSSKTLQAEVPLSDCVQKASKPPSSTQIMVKTNMYHNEKVNFHVECKDYVKKAKV 240
              |||||||||||||||||||||||| ||||||||||||||||||||||||:|||||||||||
Db        181 AAAPVCSSKTLQAEVPLSDCVQKASKPTSSTQIMVKTNMYHNEKVNFHVECRDYVKKAKV 240

Qy        241 KINPVQQSRPLLSQIHTDAAKENTCYCGAVAKRQEKKGMEPLQGHATPALPFKETQELLL 300
              ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db        241 KINPVQQSRPLLSQIHTDAAKENTCYCGAVAKRQEKKGMEPLQGHATPALPFKETQELLL 300

Qy        301 SPLPQEGPGSLAAGESSSLSASTSVSDSSQKKEEHNYSLFVSDNLGEQPTKCSPEEDEED 360
              ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db        301 SPLPQEGPGSLAAGESSSLSASTSVSDSSQKKEEHNYSLFVSDNLGEQPTKCSPEEDEED 360
```

```
Qy        361 EEDVDDEDHDEGFGSEHELSENEEEEEEEEDYEDDKDDDISDTFSEPGYENDSVEDLKEV 420
              ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db        361 EEDVDDEDHDEGFGSEHELSENEEEEEEEEDYEDDKDDDISDTFSEPGYENDSVEDLKEV 420

Qy        421 TSISSRKRGKRRYFWEYSEQLTPSQQER 448
              ||||||||||||||||||||||||||||
Db        421 TSISSRKRGKRRYFWEYSEQLTPSQQER 448


RESULT 4
Q8CDG5_MOUSE
ID   Q8CDG5_MOUSE PRELIMINARY;      PRT;   640 AA.
AC   Q8CDG5;
DT   01-MAR-2003 (TrEMBLrel. 23, Created)
DT   01-MAR-2003 (TrEMBLrel. 23, Last sequence update)
DT   01-MAR-2004 (TrEMBLrel. 26, Last annotation update)
DE   Mus musculus adult male testis cDNA, RIKEN full-length enriched
DE   library, clone:4932441F15 product:hypothetical protein, full insert
DE   sequence.
GN   Name=A930001N09Rik;
OS   Mus musculus (Mouse).
OC   Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;
OC   Mammalia; Eutheria; Euarchontoglires; Glires; Rodentia; Sciurognathi;
OC   Muroidea; Muridae; Murinae; Mus.
OX   NCBI_TaxID=10090;
RN   [1]
RP   NUCLEOTIDE SEQUENCE.
RC   STRAIN=C57BL/6J; TISSUE=Testis;
RX   MEDLINE=99279253; PubMed=10349636; DOI=10.1016/S0076-6879(99)03004-9;
RA   Carninci P., Hayashizaki Y.;
RT   "High-efficiency full-length cDNA cloning.";
RL   Meth. Enzymol. 303:19-44(1999).
RN   [2]
RP   NUCLEOTIDE SEQUENCE.
RC   STRAIN=C57BL/6J; TISSUE=Testis;
RX   MEDLINE=21085660; PubMed=11217851; DOI=10.1038/35055500;
RA   Kawai J., Shinagawa A., Shibata K., Yoshino M., Itoh M., Ishii Y.,
RA   Arakawa T., Hara A., Fukunishi Y., Konno H., Adachi J., Fukuda S.,
RA   Aizawa K., Izawa M., Nishi K., Kiyosawa H., Kondo S., Yamanaka I.,
RA   Saito T., Okazaki Y., Gojobori T., Bono H., Kasukawa T., Saito R.,
RA   Kadota K., Matsuda H.A., Ashburner M., Batalov S., Casavant T.,
RA   Fleischmann W., Gaasterland T., Gissi C., King B., Kochiwa H.,
RA   Kuehl P., Lewis S., Matsuo Y., Nikaido I., Pesole G., Quackenbush J.,
RA   Schriml L.M., Staubli F., Suzuki R., Tomita M., Wagner L., Washio T.,
RA   Sakai K., Okido T., Furuno M., Aono H., Baldarelli R., Barsh G.,
RA   Blake J., Boffelli D., Bojunga N., Carninci P., de Bonaldo M.F.,
RA   Brownstein M.J., Bult C., Fletcher C., Fujita M., Gariboldi M.,
RA   Gustincich S., Hill D., Hofmann M., Hume D.A., Kamiya M., Lee N.H.,
RA   Lyons P., Marchionni L., Mashima J., Mazzarelli J., Mombaerts P.,
RA   Nordone P., Ring B., Ringwald M., Rodriguez I., Sakamoto N.,
RA   Sasaki H., Sato K., Schoenbach C., Seya T., Shibata Y., Storch K.-F.,
RA   Suzuki H., Toyo-oka K., Wang K.H., Weitz C., Whittaker C., Wilming L.,
RA   Wynshaw-Boris A., Yoshida K., Hasegawa Y., Kawaji H., Kohtsuki S.,
RA   Hayashizaki Y.;
RT   "Functional annotation of a full-length mouse cDNA collection.";
RL   Nature 409:685-690(2001).
RN   [3]
RP   NUCLEOTIDE SEQUENCE.
RC   STRAIN=C57BL/6J; TISSUE=Testis;
RA   The FANTOM Consortium,
RA   the RIKEN Genome Exploration Research Group Phase I & II Team;
RT   "Analysis of the mouse transcriptome based on functional annotation of
RT   60,770 full-length cDNAs.";
RL   Nature 420:563-573(2002).
RN   [4]
RP   NUCLEOTIDE SEQUENCE.
RC   STRAIN=C57BL/6J; TISSUE=Testis;
RX   MEDLINE=20499374; PubMed=11042159; DOI=10.1101/gr.145100;
RA   Carninci P., Shibata Y., Hayatsu N., Sugahara Y., Shibata K., Itoh M.,
RA   Konno H., Okazaki Y., Muramatsu M., Hayashizaki Y.;
RT   "Normalization and subtraction of cap-trapper-selected cDNAs to
RT   prepare full-length cDNA libraries for rapid discovery of new genes.";
RL   Genome Res. 10:1617-1630(2000).
RN   [5]
RP   NUCLEOTIDE SEQUENCE.
RC   STRAIN=C57BL/6J; TISSUE=Testis;
RX   MEDLINE=20530913; PubMed=11076861; DOI=10.1101/gr.152600;
RA   Shibata K., Itoh M., Aizawa K., Nagaoka S., Sasaki N., Carninci P.,
RA   Konno H., Akiyama J., Nishi K., Kitsunai T., Tashiro H., Itoh M.,
RA   Sumi N., Ishii Y., Nakamura S., Hazama M., Nishine T., Harada A.,
RA   Yamamoto R., Matsumoto H., Sakaguchi S., Ikegami T., Kashiwagi K.,
RA   Fujiwake S., Inoue K., Togawa Y., Izawa M., Ohara E., Watahiki M.,
RA   Yoneda Y., Ishikawa T., Ozawa K., Tanaka T., Matsuura S., Kawai J.,
RA   Okazaki Y., Muramatsu M., Inoue Y., Kira A., Hayashizaki Y.;
RT   "RIKEN integrated sequence analysis (RISA) system-384-format
RT   sequencing pipeline with 384 multicapillary sequencer.";
RL   Genome Res. 10:1757-1771(2000).
RN   [6]
RP   NUCLEOTIDE SEQUENCE.
RC   STRAIN=C57BL/6J; TISSUE=Testis;
RA   Adachi J., Aizawa K., Akimura T., Arakawa T., Bono H., Carninci P.,
RA   Fukuda S., Furuno M., Hanagaki T., Hara A., Hashizume W.,
RA   Hayashida K., Hayatsu N., Hiramoto K., Hiraoka T., Hirozane T.,
RA   Hori F., Imotani K., Ishii Y., Itoh M., Kagawa I., Kasukawa T.,
RA   Katoh H., Kawai J., Kojima Y., Kondo S., Konno H., Kouda M., Koya S.,
RA   Kurihara C., Matsuyama T., Miyazaki A., Murata M., Nakamura M.,
RA   Nishi K., Nomura K., Numazaki R., Ohno M., Ohsato N., Okazaki Y.,
RA   Saito R., Saitoh H., Sakai C., Sakai K., Sakazume N., Sano H.,
RA   Sasaki D., Shibata K., Shinagawa A., Shiraki T., Sogabe Y., Tagami M.,
RA   Tagawa A., Takahashi F., Takaku-Akahira S., Takeda Y., Tanaka T.,
RA   Tomaru A., Toya T., Yasunishi A., Muramatsu M., Hayashizaki Y.;
RL   Submitted (JUL-2001) to the EMBL/GenBank/DDBJ databases.
DR   EMBL; AK030092; BAC26779.1; -; mRNA.
DR   Ensembl; ENSMUSG00000048249; Mus musculus.
DR   MGI; MGI:1924378; A930001N09Rik.
DR   InterPro; IPR004827; TF_bZIP.
DR   PROSITE; PS00036; BZIP_BASIC; UNKNOWN_1.
KW   Hypothetical protein.
SQ   SEQUENCE   640 AA;  72598 MW;  FE02C532FA34E1DE CRC64;

  Query Match             92.9%;  Score 2195.5;  DB 2;  Length 640;
  Best Local Similarity   92.4%;  Pred. No. 7.1e-105;
  Matches  415;  Conservative   15;  Mismatches   18;  Indels    1;  Gaps    1;

Qy          1 MPQPSVSGMDPPFGDAFRSHTFSEQTLMSTDLLANSSDPDFMYELDREMNYQQNPRDNFL 60
              ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db          1 MPQPSVSGMDPPFGDAFRSHTFSEQTLMSTDLLANSSDPDFMYELDREMNYQQNPRDNFL 60

Qy         61 SLEDCKDIENLESFTDVLDNEGALTSNWEQWDTYCEDLTKYTKLTSCDIWGTKEVDYLGL 120
              ||||||||||:|||||||| ||||||||||||||||||||||||||||||||||||||||||
Db         61 SLEDCKDIENLETFTDVLDNEDALTSNWEQWDTYCEDLTKYTKLTSCDIWGTKEVDYLGL 120

Qy        121 DDFSSPYQDEEVISKTPTLAQLNSEDSQSVSDSLYYPDSLFSVKQNPL-PSSFPGKKITS 179
              |||||||||||||||||||||||||||||||||||||||||||||||| ||||| ||||:
Db        121 DDFSSPYQDEEVISKTPTLAQLNSEDSQSVSDSLYYPDSLFSVKQNPLPPSSFPSKKITN 180

Qy        180 RAAAPVCSSKTLQAEVPLSDCVQKASKPPSSTQIMVKTNMYHNEKVNFHVECKDYVKKAK 239
              ||||||||||||||||| |||||||||| ||||||||||||||||||||||||||||||||
Db        181 RAAAPVCSSKTLQAEVPSSDCVQKASKPTSSTQIMVKTNMYHNEKVNFHVECKDYVKKAK 240

Qy        240 VKINPVQQSRPLLSQIHTDAAKENTCYCGAVAKRQEKKGMEPLQGHATPALPFKETQELL 299
              |||||||| |||||:| ||||||||||||||::|:|| ||  ||||||||||||
Db        241 VKINPVQQGRPLLSQVHIDAAKENTCYCGAVAKRQERRGVEPHQGRGTPALPFKETQELL 300

Qy        300 LSPLPQEGPGSLAAGESSSLSASTSVSDSSQKKEEHNYSLFVSDNLGEQPTKCSPEEDEE 359
              |||| |: || :|  || |||||||||||||||||||||||: ||||| |||:||:
Db        301 LSPLTQDSPGLVATAESGSLSASTSVSDSSQKKEEHNYSLFVSDNMREQPTKYSPEDDED 360

Qy        360 DEEDVDDEDHDEGFGSEHELSENEEEEEEEEDYEDDKDDDISDTFSEPGYENDSVEDLKE 419
              ||:: ||||||||||||||||||||||||||||:|||||||||||||||||||||||||||
Db        361 DEDEFDDEDHDEGFGSEHELSENEEEEEEEEDYEDDRDDDISDTFSEPGYENDSVEDLKE 420

Qy        420 VTSISSRKRGKRRYFWEYSEQLTPSQQER 448
              :|||||||||||||||||||||||||||||
Db        421 MTSISSRKRGKRRYFWEYSEQLTPSQQER 449
```

```
RESULT 5
Q5HYG4_HUMAN
ID   Q5HYG4_HUMAN PRELIMINARY;      PRT;   408 AA.
AC   Q5HYG4;
DT   10-MAY-2005 (TrEMBLrel. 30, Created)
DT   10-MAY-2005 (TrEMBLrel. 30, Last sequence update)
DT   10-MAY-2005 (TrEMBLrel. 30, Last annotation update)
DE   Hypothetical protein DKFZp686G2059 (Fragment).
GN   Name=DKFZp686G2059;
OS   Homo sapiens (Human).
OC   Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;
OC   Mammalia; Eutheria; Euarchontoglires; Primates; Catarrhini; Hominidae;
OC   Homo.
OX   NCBI_TaxID=9606;
RN   [1]
RP   NUCLEOTIDE SEQUENCE.
RC   TISSUE=Endometrium;
RG   The German cDNA Consortium;
RA   Poustka A., Albert R., Moosmayer P., Schupp I., Wellenreuther R.,
RA   Mewes H.W., Weil B., Amid C., Osanger A., Fobo G., Han M., Wiemann S.;
RL   Submitted (JAN-2005) to the EMBL/GenBank/DDBJ databases.
DR   EMBL; BX647768; CAI46039.1; -; mRNA.
KW   Hypothetical protein.
FT   NON_TER     408    408
SQ   SEQUENCE   408 AA;  45625 MW;  E3E36BFEA8B4284B CRC64;

  Query Match             90.9%;  Score 2148;  DB 2;  Length 408;
  Best Local Similarity   99.8%;  Pred. No. 1.2e-102;
  Matches  407;  Conservative    0;  Mismatches    1;  Indels    0;  Gaps    0;

Qy          1 MPQPSVSGMDPPFGDAFRSHTFSEQTLMSTDLLANSSDPDFMYELDREMNYQQNPRDNFL 60
              ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db          1 MPQPSVSGMDPPFGDAFRSHTFSEQTLMSTDLLANSSDPDFMYELDREMNYQQNPRDNFL 60

Qy         61 SLEDCKDIENLESFTDVLDNEGALTSNWEQWDTYCEDLTKYTKLTSCDIWGTKEVDYLGL 120
              ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db         61 SLEDCKDIENLESFTDVLDNEGALTSNWEQWDTYCEDLTKYTKLTSCDIWGTKEVDYLGL 120

Qy        121 DDFSSPYQDEEVISKTPTLAQLNSEDSQSVSDSLYYPDSLFSVKQNPLPSSFPGKKITSR 180
              ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db        121 DDFSSPYQDEEVISKTPTLAQLNSEDSQSVSDSLYYPDSLFSVKQNPLPSSFPGKKITSR 180

Qy        181 AAAPVCSSKTLQAEVPLSDCVQKASKPPSSTQIMVKTNMYHNEKVNFHVECKDYVKKAKV 240
              |||||||||||||||||||||||||| ||||||||||||||||||||||||||||||||
Db        181 AAAPVCSSKTLQAEVPLSDCVQKASKPTSSTQIMVKTNMYHNEKVNFHVECKDYVKKAKV 240

Qy        241 KINPVQQSRPLLSQIHTDAAKENTCYCGAVAKRQEKKGMEPLQGHATPALPFKETQELLL 300
              ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db        241 KINPVQQSRPLLSQIHTDAAKENTCYCGAVAKRQEKKGMEPLQGHATPALPFKETQELLL 300

Qy        301 SPLPQEGPGSLAAGESSSLSASTSVSDSSQKKEEHNYSLFVSDNLGEQPTKCSPEEDEED 360
              ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db        301 SPLPQEGPGSLAAGESSSLSASTSVSDSSQKKEEHNYSLFVSDNLGEQPTKCSPEEDEED 360

Qy        361 EEDVDDEDHDEGFGSEHELSENEEEEEEEEDYEDDKDDDISDTFSEPG 408
              |||||||||||||||||||||||||||||||||||||||||||||||||
Db        361 EEDVDDEDHDEGFGSEHELSENEEEEEEEEDYEDDKDDDISDTFSEPG 408


RESULT 6
Q86YR3_HUMAN
ID   Q86YR3_HUMAN PRELIMINARY;      PRT;   417 AA.
AC   Q86YR3;
DT   01-JUN-2003 (TrEMBLrel. 24, Created)
DT   01-JUN-2003 (TrEMBLrel. 24, Last sequence update)
DT   01-JUN-2003 (TrEMBLrel. 24, Last annotation update)
DE   Hypothetical protein.
OS   Homo sapiens (Human).
OC   Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;
OC   Mammalia; Eutheria; Euarchontoglires; Primates; Catarrhini; Hominidae;
OC   Homo.
OX   NCBI_TaxID=9606;
RN   [1]
RP   NUCLEOTIDE SEQUENCE.
RA   Li F., Yao K.T.;
RL   Submitted (NOV-2002) to the EMBL/GenBank/DDBJ databases.
DR   EMBL; AY174896; AAO18732.1; -; mRNA.
KW   Hypothetical protein.
SQ   SEQUENCE   417 AA;  46579 MW;  9DBD37B07C14556B CRC64;

  Query Match             90.9%;  Score 2148;  DB 2;  Length 417;
  Best Local Similarity   99.8%;  Pred. No. 1.2e-102;
  Matches  407;  Conservative    0;  Mismatches    1;  Indels    0;  Gaps    0;

Qy          1 MPQPSVSGMDPPFGDAFRSHTFSEQTLMSTDLLANSSDPDFMYELDREMNYQQNPRDNFL 60
              ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db          1 MPQPSVSGMDPPFGDAFRSHTFSEQTLMSTDLLANSSDPDFMYELDREMNYQQNPRDNFL 60

Qy         61 SLEDCKDIENLESFTDVLDNEGALTSNWEQWDTYCEDLTKYTKLTSCDIWGTKEVDYLGL 120
              ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db         61 SLEDCKDIENLESFTDVLDNEGALTSNWEQWDTYCEDLTKYTKLTSCDIWGTKEVDYLGL 120

Qy        121 DDFSSPYQDEEVISKTPTLAQLNSEDSQSVSDSLYYPDSLFSVKQNPLPSSFPGKKITSR 180
              ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db        121 DDFSSPYQDEEVISKTPTLAQLNSEDSQSVSDSLYYPDSLFSVKQNPLPSSFPGKKITSR 180

Qy        181 AAAPVCSSKTLQAEVPLSDCVQKASKPPSSTQIMVKTNMYHNEKVNFHVECKDYVKKAKV 240
              |||||||||||||||||||||||||| ||||||||||||||||||||||||||||||||
Db        181 AAAPVCSSKTLQAEVPLSDCVQKASKPTSSTQIMVKTNMYHNEKVNFHVECKDYVKKAKV 240

Qy        241 KINPVQQSRPLLSQIHTDAAKENTCYCGAVAKRQEKKGMEPLQGHATPALPFKETQELLL 300
              ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db        241 KINPVQQSRPLLSQIHTDAAKENTCYCGAVAKRQEKKGMEPLQGHATPALPFKETQELLL 300

Qy        301 SPLPQEGPGSLAAGESSSLSASTSVSDSSQKKEEHNYSLFVSDNLGEQPTKCSPEEDEED 360
              ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db        301 SPLPQEGPGSLAAGESSSLSASTSVSDSSQKKEEHNYSLFVSDNLGEQPTKCSPEEDEED 360

Qy        361 EEDVDDEDHDEGFGSEHELSENEEEEEEEEDYEDDKDDDISDTFSEPG 408
              |||||||||||||||||||||||||||||||||||||||||||||||||
Db        361 EEDVDDEDHDEGFGSEHELSENEEEEEEEEDYEDDKDDDISDTFSEPG 408


RESULT 7
Q8C8N7_MOUSE
ID   Q8C8N7_MOUSE PRELIMINARY;      PRT;   351 AA.
AC   Q8C8N7;
DT   01-MAR-2003 (TrEMBLrel. 23, Created)
DT   01-MAR-2003 (TrEMBLrel. 23, Last sequence update)
DT   01-JUN-2003 (TrEMBLrel. 24, Last annotation update)
DE   Mus musculus adult retina cDNA, RIKEN full-length enriched library,
DE   clone:A930040G19 product:hypothetical protein, full insert sequence.
DE   (Fragment).
GN   Name=A930001N09Rik;
OS   Mus musculus (Mouse).
OC   Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;
OC   Mammalia; Eutheria; Euarchontoglires; Glires; Rodentia; Sciurognathi;
OC   Muroidea; Muridae; Murinae; Mus.
OX   NCBI_TaxID=10090;
RN   [1]
RP   NUCLEOTIDE SEQUENCE.
RC   STRAIN=C57BL/6J; TISSUE=Retina;
RX   MEDLINE=99279253; PubMed=10349636; DOI=10.1016/S0076-6879(99)03004-9;
RA   Carninci P., Hayashizaki Y.;
RT   "High-efficiency full-length cDNA cloning.";
RL   Meth. Enzymol. 303:19-44(1999).
RN   [2]
RP   NUCLEOTIDE SEQUENCE.
RC   STRAIN=C57BL/6J; TISSUE=Retina;
RX   MEDLINE=21085660; PubMed=11217851; DOI=10.1038/35055500;
RA   Kawai J., Shinagawa A., Shibata K., Yoshino M., Itoh M., Ishii Y.,
RA   Arakawa T., Hara A., Fukunishi Y., Konno H., Adachi J., Fukuda S.,
RA   Aizawa K., Izawa M., Nishi K., Kiyosawa H., Kondo S., Yamanaka I.,
RA   Saito T., Okazaki Y., Gojobori T., Bono H., Kasukawa T., Saito R.,
```

```
RA   Kadota K., Matsuda H.A., Ashburner M., Batalov S., Casavant T.,
RA   Fleischmann W., Gaasterland T., Gissi C., King B., Kochiwa H.,
RA   Kuehl P., Lewis S., Matsuo Y., Nikaido I., Pesole G., Quackenbush J.,
RA   Schriml L.M., Staubli F., Suzuki R., Tomita M., Wagner L., Washio T.,
RA   Sakai K., Okido T., Furuno M., Aono H., Baldarelli R., Barsh G.,
RA   Blake J., Boffelli D., Bojunga N., Carninci P., de Bonaldo M.F.,
RA   Brownstein M.J., Bult C., Fletcher C., Fujita M., Gariboldi M.,
RA   Gustincich S., Hill D., Hofmann M., Hume D.A., Kamiya M., Lee N.H.,
RA   Lyons P., Marchionni L., Mashima J., Mazzarelli J., Mombaerts P.,
RA   Nordone P., Ring B., Ringwald M., Rodriguez I., Sakamoto N.,
RA   Sasaki H., Sato K., Schoenbach C., Seya T., Shibata Y., Storch K.-F.,
RA   Suzuki H., Toyo-oka K., Wang K.H., Weitz C., Whittaker C., Wilming L.,
RA   Wynshaw-Boris A., Yoshida K., Hasegawa Y., Kawaji H., Kohtsuki S.,
RA   Hayashizaki Y.;
RT   "Functional annotation of a full-length mouse cDNA collection.";
RL   Nature 409:685-690(2001).
RN   [3]
RP   NUCLEOTIDE SEQUENCE.
RC   STRAIN=C57BL/6J; TISSUE=Retina;
RA   The FANTOM Consortium,
RA   the RIKEN Genome Exploration Research Group Phase I & II Team;
RT   "Analysis of the mouse transcriptome based on functional annotation of
RT   60,770 full-length cDNAs.";
RL   Nature 420:563-573(2002).
RN   [4]
RP   NUCLEOTIDE SEQUENCE.
RC   STRAIN=C57BL/6J; TISSUE=Retina;
RX   MEDLINE=20499374; PubMed=11042159; DOI=10.1101/gr.145100;
RA   Carninci P., Shibata Y., Hayatsu N., Sugahara Y., Shibata K., Itoh M.,
RA   Konno H., Okazaki Y., Muramatsu M., Hayashizaki Y.;
RT   "Normalization and subtraction of cap-trapper-selected cDNAs to
RT   prepare full-length cDNA libraries for rapid discovery of new genes.";
RL   Genome Res. 10:1617-1630(2000).
RN   [5]
RP   NUCLEOTIDE SEQUENCE.
RC   STRAIN=C57BL/6J; TISSUE=Retina;
RX   MEDLINE=20530913; PubMed=11076861; DOI=10.1101/gr.152600;
RA   Shibata K., Itoh M., Aizawa K., Nagaoka S., Sasaki N., Carninci P.,
RA   Konno H., Akiyama J., Nishi K., Kitsunai T., Tashiro H., Itoh M.,
RA   Sumi N., Ishii Y., Nakamura S., Hazama M., Nishine T., Harada A.,
RA   Yamamoto R., Matsumoto H., Sakaguchi S., Ikegami T., Kashiwagi K.,
RA   Fujiwake S., Inoue K., Togawa Y., Izawa M., Ohara E., Watahiki M.,
RA   Yoneda Y., Ishikawa T., Ozawa K., Tanaka T., Matsuura S., Kawai J.,
RA   Okazaki Y., Muramatsu M., Inoue Y., Kira A., Hayashizaki Y.;
RT   "RIKEN integrated sequence analysis (RISA) system-384-format
RT   sequencing pipeline with 384 multicapillary sequencer.";
RL   Genome Res. 10:1757-1771(2000).
RN   [6]
RP   NUCLEOTIDE SEQUENCE.
RC   STRAIN=C57BL/6J; TISSUE=Retina;
RA   Adachi J., Aizawa K., Akimura T., Arakawa T., Bono H., Carninci P.,
RA   Fukuda S., Furuno M., Hanagaki T., Hara A., Hashizume W.,
RA   Hayashida K., Hayatsu N., Hiramoto K., Hiraoka T., Hirozane T.,
RA   Hori F., Imotani K., Ishii Y., Itoh M., Kagawa I., Kasukawa T.,
RA   Katoh H., Kawai J., Kojima Y., Kondo S., Konno H., Kouda M., Koya S.,
RA   Kurihara C., Matsuyama T., Miyazaki A., Murata M., Nakamura M.,
RA   Nishi K., Nomura K., Numazaki R., Ohno M., Ohsato N., Okazaki Y.,
RA   Saito R., Saitoh H., Sakai C., Sakai K., Sakazume N., Sano H.,
RA   Sasaki D., Shibata K., Shinagawa A., Shiraki T., Sogabe Y., Tagami M.,
RA   Tagawa A., Takahashi F., Takaku-Akahira S., Takeda Y., Tanaka T.,
RA   Tomaru A., Toya T., Yasunishi A., Muramatsu M., Hayashizaki Y.;
RL   Submitted (JUL-2001) to the EMBL/GenBank/DDBJ databases.
DR   EMBL; AK044760; BAC32070.1; -; mRNA.
DR   Ensembl; ENSMUSG00000048249; Mus musculus.
DR   MGI; MGI:1924378; A930001N09Rik.
KW   Hypothetical protein.
FT   NON_TER     351    351
SQ   SEQUENCE   351 AA;  39254 MW;  ABA4B36D996AFAAD CRC64;

  Query Match             72.1%;  Score 1704.5;  DB 2;  Length 351;
  Best Local Similarity   92.6%;  Pred. No. 5.9e-80;
  Matches  325;  Conservative    9;  Mismatches   16;  Indels    1;  Gaps    1;

Qy          1 MPQPSVSGMDPPFGDAFRSHTFSEQTLMSTDLLANSSDPDFMYELDREMNYQQNPRDNFL 60
              ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db          1 MPQPSVSGMDPPFGDAFRSHTFSEQTLMSTDLLANSSDPDFMYELDREMNYQQNPRDNFL 60

Qy         61 SLEDCKDIENLESFTDVLDNEGALTSNWEQWDTYCEDLTKYTKLTSCDIWGTKEVDYLGL 120
              ||||||||||:|||||||| ||||||||||||||||||||||||||||||||||||||||
Db         61 SLEDCKDIENLETFTDVLDNEDALTSNWEQWDTYCEDLTKYTKLTSCDIWGTKEVDYLGL 120

Qy        121 DDFSSPYQDEEVISKTPTLAQLNSEDSQSVSDSLYYPDSLFSVKQNPL-PSSFPGKKITS 179
              |||||||||||||||||||||||||||||||||||||||||||||||| ||||| ||||:
Db        121 DDFSSPYQDEEVISKTPTLAQLNSEDSQSVSDSLYYPDSLFSVKQNPLPPSSFPSKKITN 180

Qy        180 RAAAPVCSSKTLQAEVPLSDCVQKASKPPSSTQIMVKTNMYHNEKVNFHVECKDYVKKAK 239
              |||||||||||||||| |||||||||| ||||||||||||||||||||||||||||||||
Db        181 RAAAPVCSSKTLQAEVPSSDCVQKASKPTSSTQIMVKTNMYHNEKVNFHVECKDYVKKAK 240

Qy        240 VKINPVQQSRPLLSQIHTDAAKENTCYCGAVAKRQEKKGMEPLQGHATPALPFKETQELL 299
              |||||||| |||||:| ||||||||||||||||||::|:|| ||  ||||||||||||||
Db        241 VKINPVQQGRPLLSQVHIDAAKENTCYCGAVAKRQERRGVEPHQGRGTPALPFKETQELL 300

Qy        300 LSPLPQEGPGSLAAGESSSLSASTSVSDSSQKKEEHNYSLFVSDNLGEQPT 350
              |||| |: || :|  || |||||||||||||||||||||||||||: ||||
Db        301 LSPLTQDSPGLVATAESGSLSASTSVSDSSQKKEEHNYSLFVSDNMREQPT 351


RESULT 8
Q9CTQ7_MOUSE
ID   Q9CTQ7_MOUSE PRELIMINARY;      PRT;   330 AA.
AC   Q9CTQ7;
DT   01-JUN-2001 (TrEMBLrel. 17, Created)
DT   01-JUN-2001 (TrEMBLrel. 17, Last sequence update)
DT   01-JUN-2001 (TrEMBLrel. 17, Last annotation update)
DE   Mus musculus adult retina cDNA, RIKEN full-length enriched library,
DE   clone:A930001N09 product:hypothetical protein, full insert sequence.
DE   (Fragment).
GN   Name=A930001N09Rik;
OS   Mus musculus (Mouse).
OC   Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;
OC   Mammalia; Eutheria; Euarchontoglires; Glires; Rodentia; Sciurognathi;
OC   Muroidea; Muridae; Murinae; Mus.
OX   NCBI_TaxID=10090;
RN   [1]
RP   NUCLEOTIDE SEQUENCE.
RC   STRAIN=C57BL/6J; TISSUE=Retina;
RX   MEDLINE=99279253; PubMed=10349636; DOI=10.1016/S0076-6879(99)03004-9;
RA   Carninci P., Hayashizaki Y.;
RT   "High-efficiency full-length cDNA cloning.";
RL   Meth. Enzymol. 303:19-44(1999).
RN   [2]
RP   NUCLEOTIDE SEQUENCE.
RC   STRAIN=C57BL/6J; TISSUE=Retina;
RX   MEDLINE=21085660; PubMed=11217851; DOI=10.1038/35055500;
RA   Kawai J., Shinagawa A., Shibata K., Yoshino M., Itoh M., Ishii Y.,
RA   Arakawa T., Hara A., Fukunishi Y., Konno H., Adachi J., Fukuda S.,
RA   Aizawa K., Izawa M., Nishi K., Kiyosawa H., Kondo S., Yamanaka I.,
RA   Saito T., Okazaki Y., Gojobori T., Bono H., Kasukawa T., Saito R.,
RA   Kadota K., Matsuda H.A., Ashburner M., Batalov S., Casavant T.,
RA   Fleischmann W., Gaasterland T., Gissi C., King B., Kochiwa H.,
RA   Kuehl P., Lewis S., Matsuo Y., Nikaido I., Pesole G., Quackenbush J.,
RA   Schriml L.M., Staubli F., Suzuki R., Tomita M., Wagner L., Washio T.,
RA   Sakai K., Okido T., Furuno M., Aono H., Baldarelli R., Barsh G.,
RA   Blake J., Boffelli D., Bojunga N., Carninci P., de Bonaldo M.F.,
RA   Brownstein M.J., Bult C., Fletcher C., Fujita M., Gariboldi M.,
RA   Gustincich S., Hill D., Hofmann M., Hume D.A., Kamiya M., Lee N.H.,
RA   Lyons P., Marchionni L., Mashima J., Mazzarelli J., Mombaerts P.,
RA   Nordone P., Ring B., Ringwald M., Rodriguez I., Sakamoto N.,
RA   Sasaki H., Sato K., Schoenbach C., Seya T., Shibata Y., Storch K.-F.,
RA   Suzuki H., Toyo-oka K., Wang K.H., Weitz C., Whittaker C., Wilming L.,
RA   Wynshaw-Boris A., Yoshida K., Hasegawa Y., Kawaji H., Kohtsuki S.,
RA   Hayashizaki Y.;
RT   "Functional annotation of a full-length mouse cDNA collection.";
RL   Nature 409:685-690(2001).
RN   [3]
```

```
RP   NUCLEOTIDE SEQUENCE.
RC   STRAIN=C57BL/6J; TISSUE=Retina;
RA   The FANTOM Consortium,
RA   the RIKEN Genome Exploration Research Group Phase I & II Team;
RT   "Analysis of the mouse transcriptome based on functional annotation of
RT   60,770 full-length cDNAs.";
RL   Nature 420:563-573(2002).
RN   [4]
RP   NUCLEOTIDE SEQUENCE.
RC   STRAIN=C57BL/6J; TISSUE=Retina;
RX   MEDLINE=20499374; PubMed=11042159; DOI=10.1101/gr.145100;
RA   Carninci P., Shibata Y., Hayatsu N., Sugahara Y., Shibata K., Itoh M.,
RA   Konno H., Okazaki Y., Muramatsu M., Hayashizaki Y.;
RT   "Normalization and subtraction of cap-trapper-selected cDNAs to
RT   prepare full-length cDNA libraries for rapid discovery of new genes.";
RL   Genome Res. 10:1617-1630(2000).
RN   [5]
RP   NUCLEOTIDE SEQUENCE.
RC   STRAIN=C57BL/6J; TISSUE=Retina;
RX   MEDLINE=20530913; PubMed=11076861; DOI=10.1101/gr.152600;
RA   Shibata K., Itoh M., Aizawa K., Nagaoka S., Sasaki N., Carninci P.,
RA   Konno H., Akiyama J., Nishi K., Kitsunai T., Tashiro H., Itoh M.,
RA   Sumi N., Ishii Y., Nakamura S., Hazama M., Nishine T., Harada A.,
RA   Yamamoto R., Matsumoto H., Sakaguchi S., Ikegami T., Kashiwagi K.,
RA   Fujiwake S., Inoue K., Togawa Y., Izawa M., Ohara E., Watahiki M.,
RA   Yoneda Y., Ishikawa T., Ozawa K., Tanaka T., Matsuura S., Kawai J.,
RA   Okazaki Y., Muramatsu M., Inoue Y., Kira A., Hayashizaki Y.;
RT   "RIKEN integrated sequence analysis (RISA) system-384-format
RT   sequencing pipeline with 384 multicapillary sequencer.";
RL   Genome Res. 10:1757-1771(2000).
RN   [6]
RP   NUCLEOTIDE SEQUENCE.
RC   STRAIN=C57BL/6J; TISSUE=Retina;
RA   Adachi J., Aizawa K., Akahira S., Akimura T., Arai A., Aono H.,
RA   Arakawa T., Bono H., Carninci P., Fukuda S., Fukunishi Y., Furuno M.,
RA   Hanagaki T., Hara A., Hayatsu N., Hiramoto K., Hiraoka T., Hori F.,
RA   Imotani K., Ishii Y., Itoh M., Izawa M., Kasukawa T., Kato H.,
RA   Kawai J., Kojima Y., Konno H., Kouda M., Koya S., Kurihara C.,
RA   Matsuyama T., Miyazaki A., Nishi K., Nomura K., Numazaki R., Ohno M.,
RA   Okazaki Y., Okido T., Owa C., Saito H., Saito R., Sakai C., Sakai K.,
RA   Sano H., Sasaki D., Shibata K., Shibata Y., Shinagawa A., Shiraki T.,
RA   Sogabe Y., Suzuki H., Tagami M., Tagawa A., Takahashi F., Tanaka T.,
RA   Tejima Y., Toya T., Yamamura T., Yasunishi A., Yoshida K., Yoshino M.,
RA   Muramatsu M., Hayashizaki Y.;
RL   Submitted (AUG-2000) to the EMBL/GenBank/DDBJ databases.
DR   EMBL; AK020796; BAB32214.1; -; mRNA.
DR   Ensembl; ENSMUSG00000048249; Mus musculus.
DR   MGI; MGI:1924378; A930001N09Rik.
KW   Hypothetical protein.
FT   NON_TER     330     330
SQ   SEQUENCE   330 AA;  36692 MW;  1397555C4934A64B CRC64;

  Query Match             67.8%;  Score 1602.5;  DB 2;  Length 330;
  Best Local Similarity   92.7%;  Pred. No. 9.5e-75;
  Matches  306;  Conservative    8;  Mismatches   15;  Indels    1;  Gaps    1;

Qy          1 MPQPSVSGMDPPFGDAFRSHTFSEQTLMSTDLLANSSDPDFMYELDREMNYQQNPRDNFL 60
              ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db          1 MPQPSVSGMDPPFGDAFRSHTFSEQTLMSTDLLANSSDPDFMYELDREMNYQQNPRDNFL 60

Qy         61 SLEDCKDIENLESFTDVLDNEGALTSNWEQWDTYCEDLTKYTKLTSCDIWGTKEVDYLGL 120
              ||||||||||||:||||||| ||||||||||||||||||||||||||||||||||||||||
Db         61 SLEDCKDIENLETFTDVLDNEDALTSNWEQWDTYCEDLTKYTKLTSCDIWGTKEVDYLGL 120

Qy        121 DDFSSPYQDEEVISKTPTLAQLNSEDSQSVSDSLYYPDSLFSVKQNPL-PSSFPGKKITS 179
              |||||||||||||||||||||||||||||||||||||||||||||||| ||||| ||||:
Db        121 DDFSSPYQDEEVISKTPTLAQLNSEDSQSVSDSLYYPDSLFSVKQNPLPPSSFPSKKITN 180

Qy        180 RAAAPVCSSKTLQAEVPLSDCVQKASKPPSSTQIMVKTNMYHNEKVNFHVECKDYVKKAK 239
              |||||||||||||||| |||||||||| ||||||||||||||||||||||||||||||||
Db        181 RAAAPVCSSKTLQAEVPSSDCVQKASKPTSSTQIMVKTNMYHNEKVNFHVECKDYVKKAK 240

Qy        240 VKINPVQQSRPLLSQIHTDAAKENTCYCGAVAKRQEKKGMEPLQGHATPALPFKETQELL 299
              ||||||| |||||:| ||||||||||||||||||::|:|| ||  ||||||||||||
Db        241 VKINPVQQGRPLLSQVHIDAAKENTCYCGAVAKRQERRGVEPHQGRGTPALPFKETQELL 300

Qy        300 LSPLPQEGPGSLAAGESSSLSASTSVSDSS 329
              |||| |: || :|  || ||||||||||||
Db        301 LSPLTQDSPGLVATAESGSLSASTSVSDSS 330


RESULT 9
Q5ZMD5_CHICK
ID   Q5ZMD5_CHICK PRELIMINARY;      PRT;   131 AA.
AC   Q5ZMD5;
DT   25-OCT-2004 (TrEMBLrel. 28, Created)
DT   25-OCT-2004 (TrEMBLrel. 28, Last sequence update)
DT   25-OCT-2004 (TrEMBLrel. 28, Last annotation update)
DE   Hypothetical protein.
GN   ORFNames=RCJMB04_2h14;
OS   Gallus gallus (Chicken).
OC   Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;
OC   Archosauria; Aves; Neognathae; Galliformes; Phasianidae; Phasianinae;
OC   Gallus.
OX   NCBI_TaxID=9031;
RN   [1]
RP   NUCLEOTIDE SEQUENCE.
RC   STRAIN=CB; TISSUE=Bursa;
RA   Caldwell R.B., Kierzek A.M., Arakawa H., Bezzubov Y., Zaim J.,
RA   Fiedler P., Kutter S., Blagodatski A., Kostovska D., Koter M.,
RA   Plachy J., Carninci P., Hayashizaki Y., Buerstedde J.M.;
RT   "Full-length cDNAs from chicken bursal lymphocytes to facilitate
RT   genefunction analysis.";
RL   Genome Biol. 6:R6-R6(2005).
DR   EMBL; AJ719449; CAG31108.1; -; mRNA.
KW   Hypothetical protein.
SQ   SEQUENCE   131 AA;  15183 MW;  4D73D84C28DE3670 CRC64;

  Query Match             27.7%;  Score 656;  DB 2;  Length 131;
  Best Local Similarity   90.1%;  Pred. No. 1.3e-26;
  Matches  118;  Conservative    6;  Mismatches    7;  Indels    0;  Gaps    0;

Qy          1 MPQPSVSGMDPPFGDAFRSHTFSEQTLMSTDLLANSSDPDFMYELDREMNYQQNPRDNFL 60
              |||||||||||||||||| ||||||||||||:|||||||||||||:|||: ||| |
Db          1 MPQPSVSGMDPPFGDAFRSHVFSEQTLMSTDLLASSSDPDFMYELDREMDYQQSSRDNLL 60

Qy         61 SLEDCKDIENLESFTDVLDNEGALTSNWEQWDTYCEDLTKYTKLTSCDIWGTKEVDYLGL 120
              |:|||||:||||||||:|| | | || ||||||||||||||||||||||||||||||||
Db         61 SMEDCKDLENLESFTDILDKEAAFTSKWEQWDTYCEDLTKYTKLTSCDIWGTKEVDYLGL 120

Qy        121 DDFSSPYQDEE 131
              ||||||||||
Db        121 DDFSSPYQDEE 131


RESULT 10
Q4RRX3_TETNG
ID   Q4RRX3_TETNG PRELIMINARY;      PRT;   600 AA.
AC   Q4RRX3;
DT   13-SEP-2005 (TrEMBLrel. 31, Created)
DT   13-SEP-2005 (TrEMBLrel. 31, Last sequence update)
DT   13-SEP-2005 (TrEMBLrel. 31, Last annotation update)
DE   Chromosome 7 SCAF15001, whole genome shotgun sequence.
DE   (Fragment).
GN   ORFNames=GSTENG00029962001;
OS   Tetraodon nigroviridis (Green puffer).
OC   Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;
OC   Actinopterygii; Neopterygii; Teleostei; Euteleostei; Neoteleostei;
OC   Acanthomorpha; Acanthopterygii; Percomorpha; Tetraodontiformes;
OC   Tetradontoidea; Tetraodontidae; Tetraodon.
OX   NCBI_TaxID=99883;
RN   [1]
RP   NUCLEOTIDE SEQUENCE.
RA   Jaillon O., Aury J.M., Brunet F., Petit J.L., Stange-Thomann N.,
RA   Mauceli E., Bouneau L., Fischer C., Ozouf-Costaz C., Bernot A.,
RA   Nicaud S., Jaffe D., Fisher S., Lutfalla G., Dossat C., Segurens B.,
```

```
RA   Dasilva C., Salanoubat M., Levy M., Boudet N., Castellano S.,
RA   Anthouard V., Jubin C., Castelli V., Katinka M., Vacherie B.,
RA   Biemont C., Skalli Z., Cattolico L., Poulain J., De Berardinis V.,
RA   Cruaud C., Duprat S., Brottier P., Coutanceau J.P., Gouzy J.,
RA   Parra G., Lardier G., Chapple C., McKernan K.J., McEwan P., Bosak S.,
RA   Kellis M., Volff JN., Guigo R., Zody M.C., Mesirov J.,
RA   Lindblad-Toh K., Birren B., Nusbaum C., Kahn D., Robinson-Rechavi M.,
RA   Laudet V., Schachter V., Quetier F., Saurin W., Scarpelli C.,
RA   Wincker P., Lander E.S., Weissenbach J., Roest Crollius H.;
RT   "Genome duplication in the teleost fish Tetraodon nigroviridis reveals
RT   the early vertebrate proto-karyotype.";
RL   Nature 431:946-957(2004).
RN   [2]
RP   NUCLEOTIDE SEQUENCE.
RG   Genoscope; Whitehead Institute Centre for Genome Research;
RL   Submitted (FEB-2004) to the EMBL/GenBank/DDBJ databases.
CC   -!- CAUTION: The sequence shown here is derived from an
CC       EMBL/GenBank/DDBJ whole genome shotgun (WGS) entry which is
CC       preliminary data.
DR   EMBL; CAAE01015001; CAG08859.1; -; Genomic_DNA.
FT   NON_TER       1      1
FT   NON_TER     600    600
SQ   SEQUENCE   600 AA;  66291 MW;  B2F28B2F3B27B1C4 CRC64;

  Query Match             27.2%;  Score 644;  DB 2;  Length 600;
  Best Local Similarity   34.0%;  Pred. No. 3.2e-25;
  Matches  166;  Conservative   56;  Mismatches  104;  Indels  162;  Gaps   13;

Qy          3 QPSVSGMDPPFGDAFRSHTFSEQTLMSTDLLANSSDPDFMYELDREMNYQQNP------- 55
              |||||||:|||||||::::| :| | ||:||| ||||||||||||: |:||:|
Db          1 QPSVSGMEPPFGDAFQNYSFVDQALTSTELLATSSDPDFMYELDRDTNHQQSPCGDSMVG 60

Qy         56 -----------RDNFLSLEDCKDIENLESFTDVLDNEGALTSNWEQWDTYCEDLTKYTKL 104
                        |  : | :|: : :  :|               ||||:| ||||:||:|
Db         61 VGDGGKEVEGCVDQLMGLGECETVHSSSAF-------------EQWDSYWEDLTRYTRL 106

Qy        105 TSCDIWGTKEVDYLGLDDFSSPYQDEEVISKTPTLAQLNSEDSQSVSDSLYYPDSLFSVK 164
              |||||||||||:|||||||||||||||| :|||||||||||  | ::|| |  |
Db        107 ASCDIWGTKEVDFLGLDDFSSPYQDEEVIGRTPTLAQLNSEDSLPVCEALYPPADL---- 162

Qy        165 QNPLPSSFPGKKITSRAAAPVCSSK----------------TLQAEVPLSDCV---QKA 204
                  | || :  |:    | ||                :  :  || |     |||
Db        163 ------SLPGPQPQSQPPQLPCHSKRPVVPGSVRSSSSSTSSSRPSRSPLPDFAEGSQKA 216

Qy        205 SKP-PSSTQIMVKTNMYHNEKVNFHVECKDYVK-KAKVKINPVQQSRPLLSQIHTDAAKE 262
              ::| ||||: | |          |     :|| : : |::    : : |  :  ::|
Db        217 TRPVPSSTETMAKNQ-------NLFSLSQDYGQGQPKLQARATKMAAP--APHNSDFVPP 267

Qy        263 NTCYCGAVAKRQEKKGMEPLQGHATPALP-FKETQELLLSPLPQEGPGSLAAGESSSLSA 321
              :  :  :| |  |  :   : |:|   |  :|
Db        268 ASSNLMSFERRVEATGRTDVPAGWSSAVPHLVEANRVL---------------------- 305

Qy        322 STSVSDSSQKKEEHNYSLFVSDNLGEQPTKCSPEEDEEDEEDVDDEDHDEGFGSEHELSE 381
                      | | |   |:::|     |    | |
Db        306 -----------EVNASGLASESVGAAEAICGGSCDME---------------------- 331

Qy        382 NEEEEEEEEDYEDDKDDDISDTFSEPGYENDSVEDLKEVTS-ISSRKRGKRRYFWEYSEQ 440
                                          :||:| :|: :|||||||||||||||||
Db        332 ----------------------------LMEDIKGLTAGVSSRKRGKRRYFWEYSEQ 360

Qy        441 LTPSQQER 448
                | ||:|||
Db        361 LPPSKQER 368


RESULT 11
Q9DUN0_HHV8
ID   Q9DUN0_HHV8 PRELIMINARY;      PRT;   976 AA.
AC   Q9DUN0;
DT   01-MAR-2001 (TrEMBLrel. 16, Created)
DT   01-MAR-2001 (TrEMBLrel. 16, Last sequence update)
DT   01-MAR-2004 (TrEMBLrel. 26, Last annotation update)
DE   Orf73.
OS   Human herpesvirus 8 (HHV-8) (Kaposi's sarcoma-associated herpesvirus).
OC   Viruses; dsDNA viruses, no RNA stage; Herpesviridae;
OC   Gammaherpesvirinae; Rhadinovirus.
OX   NCBI_TaxID=37296;
RN   [1]
RP   NUCLEOTIDE SEQUENCE.
RC   STRAIN=PK-1;
RX   MEDLINE=99445611; PubMed=10515805; DOI=10.1086/315098;
RA   Gao S.J., Zhang Y.J., Deng J.H., Rabkin C.S., Flore O., Jenson H.B.;
RT   "Molecular polymorphism of Kaposi's sarcoma-associated herpesvirus
RT   (Human herpesvirus 8) latent nuclear antigen: evidence for a large
RT   repertoire of viral genotypes and dual infection with different viral
RT   genotypes.";
RL   J. Infect. Dis. 180:1466-1476(1999).
RN   [2]
RP   NUCLEOTIDE SEQUENCE.
RC   STRAIN=PK-1;
RX   MEDLINE=20381179; PubMed=10900044;
RA   Zhang Y.J., Deng J.H., Rabkin C., Gao S.J.;
RT   "Hot-spot variations of Kaposi's sarcoma-associated herpesvirus latent
RT   nuclear antigen and application in genotyping by PCR-RFLP.";
RL   J. Gen. Virol. 81:2049-2058(2000).
DR   EMBL; AF192756; AAG01636.1; -; Genomic_DNA.
DR   GO; GO:0005856; C:cytoskeleton; IEA.
DR   InterPro; IPR002017; Spectrin.
SQ   SEQUENCE   976 AA;  112017 MW;  E5781E2A509FF70B CRC64;

  Query Match              8.2%;  Score 194;  DB 2;  Length 976;
  Best Local Similarity   24.0%;  Pred. No. 0.074;
  Matches   88;  Conservative   56;  Mismatches  149;  Indels   74;  Gaps   12;

Qy        124 SSPYQDEEVISKTPTLAQLNSEDSQSVSDSLYYPDSLFSVKQNPLPSSFPGKKITSRAAA 183
              |||       :|   |        |: | | :|    |  | :: || ||  |:  :|
Db        111 SSPIPPSHPVSPGTT-------DTHSPSPALPPTQSPESSQRPPL-SSPTGRPDSSTPMR 162

Qy        184 PVCSSKTLQAEVPLSDCVQKASK------PPSSTQIMVKTNMYHNEKVNFHVECKDYVKK 237
              |  | :|    | :   :  ||       ||: : : |  :                 :
Db        163 PPPSQQTTPPHSPTTPPPEPPSKSSPDSLAPSTLRSLRKRRLSS--------------PQ 208

Qy        238 AKVKINPVQQSRPLL--------SQIHTDAAKENTCYCGAVAKRQEKKGMEPLQ----GH 285
                :||: || |:           ::   | |: | |: :       |       |
Db        209 GPSTLNPICQSPPVSPPRCDFANRSVYPPWATESPIYVGSSSDGDTPPRQPPTSPISIGS 268

Qy        286 ATPALPF--KETQELLL--------SPLPQEGPGSLAAGESS-----SLSASTSVSDSSQ 330
              ::|:      :|  |:|         |   :|   :  ||| :     | :     |:
Db        269 SSPSEGSWGDDTAMLVLLAEIAEEASKNEKECSENNQAGEDNGDNEISKESQVDKDDNDN 328

Qy        331 KKEEHNYSLFVSDNLGEQPTKCSPEEDEEDEEDVDDEDHDEGFGSEHELSENEEEEEEEE 390
              | :|       :|   |:  :   |||:|:::: |||: ||    | :  |:|||:||||
Db        329 KDDEEEQE---TDEEDEEDDEEDDEEDDEEDDEEDDEEDDEEDDEEDEEEDEEEDEEEE 385

Qy        391 DYEDDKDDDISDTFSEPGYENDSVEDLKE----------VTSISSRKRGKRRYFWEYSEQ 440
              | |:::|::  |        |:|  || ||            || | :: :       :|
Db        386 DEEEEEDEEDDDDEDNEDEEDDEEEDKKEDEEDGGDGNKTLSIQSSQQQQE------PQQ 439

Qy        441 LTPSQQE 447
                | |||
Db        440 QEPQQQE 446


RESULT 12
Q91LX9_HHV8
ID   Q91LX9_HHV8 PRELIMINARY;      PRT;  1003 AA.
AC   Q91LX9;
DT   01-DEC-2001 (TrEMBLrel. 19, Created)
DT   01-DEC-2001 (TrEMBLrel. 19, Last sequence update)
DT   01-MAR-2004 (TrEMBLrel. 26, Last annotation update)
DE   ORF73.
OS   Human herpesvirus 8 (HHV-8) (Kaposi's sarcoma-associated herpesvirus).
OC   Viruses; dsDNA viruses, no RNA stage; Herpesviridae;
OC   Gammaherpesvirinae; Rhadinovirus.
OX   NCBI_TaxID=37296;
RN   [1]
```

```
RP   NUCLEOTIDE SEQUENCE.
RX   MEDLINE=21376412; PubMed=11483733;
RX   DOI=10.1128/JVI.75.17.7882-7892.2001;
RA   Garber A.C., Shu M.A., Hu J., Renne R.;
RT   "DNA binding and modulation of gene expression by the latency-
RT   associated nuclear antigen of Kaposi's sarcoma-associated
RT   herpesvirus.";
RL   J. Virol. 75:7882-7892(2001).
DR   EMBL; AF360120; AAK50002.1; -; Genomic_DNA.
DR   GO; GO:0005856; C:cytoskeleton; IEA.
DR   InterPro; IPR002017; Spectrin.
SQ   SEQUENCE   1003 AA;  115517 MW;  C20C43308B01A0A3 CRC64;

  Query Match             8.1%;  Score 191.5;  DB 2;  Length 1003;
  Best Local Similarity   24.7%;  Pred. No. 0.1;
  Matches   92;  Conservative   53;  Mismatches  139;  Indels   89;  Gaps   14;

Qy        124 SSPYQDEEVISKTPTLAQLNSEDSQSVSDSLYYPDSLFSVKQNPLPSSFPGKKITSRAAA 183
              |||      :|   |       |: | | :|    |  | :: || ||  |:  :|
Db        111 SSPIPPSHPVSPGTT-------DTHSPSPALPPTQSPESSQRPPL-SSPTGRPDSSTPMR 162

Qy        184 PVCSSKTLQAEVPLSDCVQKASK------PPSSTQIMVKTNMYHNEKVNFHVECKDYVKK 237
              |  | :|     | :   :  ||       ||: : : |  :                :
Db        163 PPPSQQTTPPHSPTTPPPEPPSKSSPDSLAPSTLRSLRKRRLSS-------------PQ 208

Qy        238 AKVKINPVQQSRPLL--------SQIHTDAAKENTCYCGAVAKRQEKKGMEPLQ----GH 285
                :||: || |:           ::   | |:  | |: :          |      |
Db        209 GPSTLNPICQSPPVSPPRCDFANRSVYPPWATESPIYVGSSSDGDTPPRQPPTSPISIGS 268

Qy        286 ATPALPF--KETQELLL--------SPLPQEGPGSLAAGESSSLSASTSVSDSSQKKEEH 335
              ::|:      :|  |:|        |   :|   :  |||    :    :|  ||  ::
Db        269 SSPSEGSWGDDTAMLVLLAEIAEEASKNEKECSENNQAGED---NGDNEISKESQVDKDD 325

Qy        336 NYSLFVSDNLGEQPTKCSPEEDEEDEE--------------DVDDEDHDEGFGSEHELS 380
              |      ||  ::  : : ||||||:|                : |||: ||    | :
Db        326 N------DNKDDEEEQETDEEDEEDDEEDDEEDDEEDDEEDDEEDDEEDDEEDEEDEE 379

Qy        381 ENEEEEEEEEDYEDDKDDDISDTFSEPGYENDSVEDLKE------VTSISSRKRGKRRYF 434
              |:||||:|||: ||::|||  |   |   | |  || ::        || | :: :
Db        380 EDEEEEDEEEEEEDEEDDDDEDNEDE---EEDKKEDEEDGGDGNKTLSIQSSQQQQE--- 433

Qy        435 WEYSEQLTPSQQE 447
                :|  | |||
Db        434 ---PQQQEPQQQE 443


RESULT 13
Q9QR71_HHV8
ID   Q9QR71_HHV8 PRELIMINARY;      PRT;  1129 AA.
AC   Q9QR71;
DT   01-MAY-2000 (TrEMBLrel. 13, Created)
DT   01-MAY-2000 (TrEMBLrel. 13, Last sequence update)
DT   01-MAR-2004 (TrEMBLrel. 26, Last annotation update)
DE   Latent nuclear antigen.
GN   Name=ORF 73;
OS   Human herpesvirus 8 (HHV-8) (Kaposi's sarcoma-associated herpesvirus).
OC   Viruses; dsDNA viruses, no RNA stage; Herpesviridae;
OC   Gammaherpesvirinae; Rhadinovirus.
OX   NCBI_TaxID=37296;
RN   [1]
RP   NUCLEOTIDE SEQUENCE.
RX   MEDLINE=99329221; PubMed=10400794;
RA   Glenn M., Rainbow L., Aurad F., Davison A., Schulz T.F.;
RT   "Identification of a spliced gene from Kaposi's sarcoma-associated
RT   herpesvirus encoding a protein with similarities to latent membrane
RT   proteins 1 and 2A of Epstein-Barr virus.";
RL   J. Virol. 73:6953-6963(1999).
RN   [2]
RP   NUCLEOTIDE SEQUENCE.
RA   Glenn M.A., Rainbow L., Aurade F., Davison A.J., Schulz T.F.;
RL   Submitted (APR-1999) to the EMBL/GenBank/DDBJ databases.
DR   EMBL; AF148805; AAD46501.1; -; Genomic_DNA.
DR   GO; GO:0005856; C:cytoskeleton; IEA.
DR   InterPro; IPR002017; Spectrin.
SQ   SEQUENCE   1129 AA;  131346 MW;  8F63855B45F79109 CRC64;

  Query Match             8.1%;  Score 191.5;  DB 2;  Length 1129;
  Best Local Similarity   25.0%;  Pred. No. 0.12;
  Matches   93;  Conservative   52;  Mismatches  142;  Indels   85;  Gaps   14;

Qy        124 SSPYQDEEVISKTPTLAQLNSEDSQSVSDSLYYPDSLFSVKQNPLPSSFPGKKITSRAAA 183
              |||      :|   |       |: | | :|    |  | :: || ||  |:  :|
Db        111 SSPIPPSHPVSPGTT-------DTHSPSPALPPTQSPESSQRPPL-SSPTGRPDSSTPMR 162

Qy        184 PVCSSKTLQAEVPLSDCVQKASK------PPSSTQIMVKTNMYHNEKVNFHVECKDYVKK 237
              |  | :|     | :   :  ||       ||: : : |  :                :
Db        163 PPPSQQTTPPHSPTTPPPEPPSKSSPDSLAPSTLRSLRKRRLSS-------------PQ 208

Qy        238 AKVKINPVQQSRPLL--------SQIHTDAAKENTCYCGAVAKRQEKKGMEPLQ----GH 285
                :||: || |:           ::   | |:  | |: :          |      |
Db        209 GPSTLNPICQSPPVSPPRCDFANRSVYPPWATESPIYVGSSSDGDTPPRQPPTSPISIGS 268

Qy        286 ATPALPF--KETQELLL--------SPLPQEGPGSLAAGESSSLSASTSVSDSSQKKEEH 335
              ::|:      :|  |:|        |   :|   :  |||    :    :|  ||  ::
Db        269 SSPSEGSWGDDTAMLVLLAEIAEEASKNEKECSENNQAGED---NGDNEISKESQVDKDD 325

Qy        336 NYSLFVSDNLGEQPTKCSPEEDEEDEEDVDDEDHDEGFGSE----------HELSENEEE 385
              | :    |:  || |    |||:|:::: |||: ||    |           |  | |:|
Db        326 NDN---KDDEEEQETDEEDEEDDEEDDEEDDEEDDEEDDEEDDEEDDEEDEEEDEEEDE 382

Qy        386 EEEEEDYEDDKDDDISDTFSEPGYENDSVEDLKE----------VTSISSRKRGKRRYFW 435
              ||:||: ||::|||  |   |   |:|  || ||            || | :: :
Db        383 EEDEEEEEDEEDDDDEDNEDE---EDDEEEDKKEDEEDGGDGNKTLSIQSSQQQQE---- 435

Qy        436 EYSEQLTPSQQE 447
                :|  | |||
Db        436 --PQQQEPQQQE 445


RESULT 14
Q7RAS7_PLAYO
ID   Q7RAS7_PLAYO PRELIMINARY;      PRT;  2649 AA.
AC   Q7RAS7;
DT   01-MAR-2004 (TrEMBLrel. 26, Created)
DT   01-MAR-2004 (TrEMBLrel. 26, Last sequence update)
DT   01-MAR-2004 (TrEMBLrel. 26, Last annotation update)
DE   Hypothetical protein.
GN   Name=PY06422;
OS   Plasmodium yoelii yoelii.
OC   Eukaryota; Alveolata; Apicomplexa; Haemosporida; Plasmodium.
OX   NCBI_TaxID=73239;
RN   [1]
RP   NUCLEOTIDE SEQUENCE.
RC   STRAIN=17XNL;
RX   MEDLINE=22255706; PubMed=12368865; DOI=10.1038/nature01099;
RA   Carlton J.M., Angiuoli S.V., Suh B.B., Kooij T.W., Pertea M.,
RA   Silva J.C., Ermolaeva M.D., Allen J.E., Selengut J.D., Koo H.L.,
RA   Peterson J.D., Pop M., Kosack D.S., Shumway M.F., Bidwell S.L.,
RA   Shallom S.J., van Aken S.E., Riedmuller S.B., Feldblyum T.V.,
RA   Cho J.K., Quackenbush J., Sedegah M., Shoaibi A., Cummings L.M.,
RA   Florens L., Yates J.R. III, Raine J.D., Sinden R.E., Harris M.A.,
RA   Cunningham D.A., Preiser P.R., Bergman L.W., Vaidya A.B.,
RA   van Lin L.H., Janse C.J., Waters A.P., Smith H.O., White O.R.,
RA   Salzberg S.L., Venter J.C., Fraser C.M., Hoffman S.L., Gardner M.J.,
RA   Carucci D.J.;
RT   "Genome sequence and comparative analysis of the model rodent malaria
RT   parasite Plasmodium yoelii yoelii.";
RL   Nature 419:512-519(2002).
CC   -!- CAUTION: The sequence shown here is derived from an
CC       EMBL/GenBank/DDBJ whole genome shotgun (WGS) entry which is
CC       preliminary data.
DR   EMBL; AABL01002172; EAA18637.1; -; Genomic_DNA.
DR   InterPro; IPR011591; Botulinum.
DR   ProDom; PD001963; Botulinum; 2.
KW   Hypothetical protein.
SQ   SEQUENCE   2649 AA;  309810 MW;  E7207F344643AC24 CRC64;
```

```
  Query Match             8.1%;  Score 190.5;  DB 2;  Length 2649;
  Best Local Similarity   20.6%;  Pred. No. 0.36;
  Matches   93;  Conservative   74;  Mismatches  116;  Indels  169;  Gaps   18;

Qy         15 DAFRSHTFSE-----QTLMSTDLLANSSDPDFMY----------ELDREM--------N 50
              |: : : | :     : |:: :   |||: |::|           ||:  |          |
Db        140 DSAKIYPFDKINNNVKNLLTNNRNDNSSNEDYIYMSNNLNYNFKEELENNMFNNLNMFKN 199

Qy         51 YQQNPRD----NFLSLEDCKDIENLESFTDVLDNEGALTSNWEQWDTYCEDLTKYTKLTS 106
                ::|  |    :|:  :  |||  |    |    || |:  :  : :   | ||: : ||:
Db        200 LKKNNFDDATKHFMKPKTFKDINGLYYDT----NEKAIDIDLLKMEH--EKLTRMSSLTN 253

Qy        107 CDIWGTKEVDYLGLDDFSSPYQDEEVISKTPTLAQLNSEDSQSVSDSLYYPDSLFSVKQN 166
               :     | |||  ::| : | | |         :::|   : : :           :::|
Db        254 LN---RVECDYLAHNNFPNAYNDNEY-----KKSKVNKRGNNNYT----------RLRKN 295

Qy        167 PLPSSFPGKKITSRAAAPVCSSKTLQAEVPLSDCVQKASKPPSSTQIMVKTNMYHNEKVN 226
               |  :    :|                              : | ||  |   :::|:    |
Db        296 ILDKNIRSDEIN-------------------------NTPLSSDDI--TSHIYNRNVEN 327

Qy        227 FHVECKDYVKKAKVKINPVQQSRPLLSQIHTDAAKENTCYCGAVAKRQEKKGMEPLQGHA 286
               ::          | :|   ::  |    :|   :  :||         : || |:  : :
Db        328 YN---------AGLKNEKIENIR----KIGNKSKGKNTILKKIGVHWENKKEMQNEESY- 373

Qy        287 TPALPFKETQELLLSPLPQEGPGSLAAGESSSLSASTSVSDSSQKKEEHNYSLFVSDNLG 346
                                                  | .:|:||
Db        374 ---------------------------------------DEEEKEEEE---------- 382

Qy        347 EQPTKCSPEEDEEDEEDVDDEDHDEGFGSEHELSENEEEE-EEEEDYEDDKDDDISDTFS 405
                     |:||::|:| |||| ::| | | |  | |||| ||||| |:::::|
Db        383 ------VEEDDEDEEDDEDDEDDEDGEGEEDEEEEEEEEEGEEEEDEEEEEEED------ 430

Qy        406 EPGYENDSVEDLKEVTSISSRKR-----GKRR 432
                  |  |: :||    |:|:     ||:|
Db        431 ------DEEEEEEEVEEYESKKKHTYKSGKKR 456
```

RESULT 15
Q98148_HHV8
ID   Q98148_HHV8 PRELIMINARY;   PRT; 1162 AA.
AC   Q98148;
DT   01-FEB-1997 (TrEMBLrel. 02, Created)
DT   01-FEB-1997 (TrEMBLrel. 02, Last sequence update)
DT   13-SEP-2005 (TrEMBLrel. 31, Last annotation update)
DE   ORF 73.
OS   Human herpesvirus 8 (HHV-8) (Kaposi's sarcoma-associated herpesvirus).
OC   Viruses; dsDNA viruses, no RNA stage; Herpesviridae;
OC   Gammaherpesvirinae; Rhadinovirus.
OX   NCBI_TaxID=37296;
RN   [1]
RP   NUCLEOTIDE SEQUENCE.
RX   MEDLINE=97048116; PubMed=8892957;
RA   Cesarman E., Nador R.G., Bai F., Bohenzky R.A., Russo J.J.,
RA   Moore P.S., Chang Y., Knowles D.M.;
RT   "Kaposi's sarcoma-associated herpesvirus contains G protein-coupled
RT   receptor and cyclin D homologs which are expressed in Kaposi's sarcoma
RT   and malignant lymphoma.";
RL   J. Virol. 70:8218-8223(1996).
RN   [2]
RP   NUCLEOTIDE SEQUENCE.
RX   MEDLINE=97094384; PubMed=8939871; DOI=10.1126/science.274.5293.1739;
RA   Moore P.S., Bashoff C., Weiss R.A., Chang Y.;
RT   "Molecular mimicry of human cytokine and cytokine response pathway
RT   genes by KSHV.";
RL   Science 274:1739-1744(1996).
RN   [3]
RP   NUCLEOTIDE SEQUENCE.
RX   MEDLINE=97121480; PubMed=8962146; DOI=10.1073/pnas.93.25.14862;
RA   Russo J.J., Bohenzky R.A., Chien M.C., Chen J., Yan M., Maddalena D.,
RA   Parry J.P., Peruzzi D., Edelman I.S., Chang Y., Moore P.S.;
RT   "Nucleotide sequence of the Kaposi sarcoma-associated herpesvirus
RT   (HHV8).";
RL   Proc. Natl. Acad. Sci. U.S.A. 93:14862-14867(1996).
RN   [4]
RP   NUCLEOTIDE SEQUENCE.
RA   Russo J.J., Bohenzky R.A., Chien M.-C., Chen J., Yan M., Maddalena D.,
RA   Parry J.P., Peruzzi D., Edelman I.S., Chang Y., Moore P.S.;
RL   Submitted (MAY-1997) to the EMBL/GenBank/DDBJ databases.
DR   EMBL; U52064; AAC55944.1; -; Genomic_DNA.
DR   EMBL; U75698; AAC57158.1; -; Genomic_DNA.
DR   InterPro; IPR002017; Spectrin.
SQ   SEQUENCE   1162 AA; 135213 MW; 1A72CE01C1CB081C CRC64;

```
  Query Match             8.0%;  Score 190;  DB 2;  Length 1162;
  Best Local Similarity   24.2%;  Pred. No. 0.15;
  Matches   94;  Conservative   50;  Mismatches  137;  Indels  108;  Gaps   14;

Qy        124 SSPYQDEEVISKTPTLAQLNSEDSQSVSDSLYYPDSLFSVKQNPLPSSPPGKKITSRAAA 183
              |||       :|   |       |: | | :|    |  | :: || ||  |:  :|
Db        111 SSPIPPSHPVSPGTT-------DTHSPSPALPPTQSPESSQRPPL-SSPTGRPDSSTPMR 162

Qy        184 PVCSSKTLQAEVPLSDCVQKASK------PPSSTQIMVKTNMYHNEKVNFHVECKDYVKK 237
              |  | :|     | :   :  ||       ||: : : |  :                 :
Db        163 PPPSQQTTPPHSPTTPPPEPPSKSSPDSLAPSTLRSLRKRRLSS-------------PQ 208

Qy        238 AKVKINPVQQSRPLL--------SQIHTDAAKENTCYCGAVAKRQEKKGMEPLQ----GH 285
                :||: || |:           ::    | |:  | |: :           |      |
Db        209 GPSTLNPICQSPPVSPPRCDFANRSVYPPWATESPIYVGSSSDGDTPPRQPPTSPISIGS 268

Qy        286 ATPALPF--KETQELLL--------SPLPQEGPGSLAAGESSSLSASTSVSDSSQKKEEH 335
              ::|:      :|  |:|        |   :|   :  |||    :   :|  ||  ::
Db        269 SSPSEGSWGDDTAMLVLLAEIAEEASKNEKECSENNQAGED---NGDNEISKESQVDKDD 325

Qy        336 NYSLFVSDNLGEQPTKCSPEEDEEDEE-------------------------DVDDED 368
              |       ||  ::  : : ||||||:|                         : |||:
Db        326 N------DNKDDEEEQETDEEDEEDDEEDDEEDDEEDDEEDDEEDDEEDDEEDDEEDEE 379

Qy        369 HDEGFGSEHELSENEEEEEEEEDYEDDKDDDISDTFSEPGYENDSVEDLKE--------- 419
               ||    | :  |:|||:||||| ||| |:|  |        |:|  || ||
Db        380 DDEEDDEEEDEEEDEEEDEEEEDEEDDDDEDNEDE------EDDEEEDKKEDEEDGGDGN 433

Qy        420 -VTSISSRKRGKRRYFWEYSEQLTPSQQE 447
                 || | :: :        :|  | |||
Db        434 KTLSIQSSQQQQE------PQQQEPQQQE 456
```

Search completed: February 28, 2006, 08:53:02
Job time : 166.304 secs

GenCore version 5.1.7

OM protein - protein search, using sw model

Run on: February 28, 2006, 08:37:25 ; Search time 0.685046 Seconds
(without alignments)
1282.772 Million cell updates/sec

Title: US-10-717-665A-44_COPY_448_449
Perfect score: 10
Sequence: 1 RM 2

Scoring table: BLOSUM62
Gapop 10.0 , Gapext 0.5

Searched: 2443163 seqs, 439378781 residues

Total number of hits satisfying chosen parameters: 2443163

Minimum DB seq length: 0
Maximum DB seq length: 2000000000

Post-processing: Minimum Match 0%
Maximum Match 100%
Listing first 45 summaries

Database : A_Geneseq_21:*
1: geneseqp1980s:*
2: geneseqp1990s:*
3: geneseqp2000s:*
4: geneseqp2001s:*
5: geneseqp2002s:*
6: geneseqp2003as:*
7: geneseqp2003bs:*
8: geneseqp2004s:*
9: geneseqp2005s:*

Pred. No. is the number of results predicted by chance to have a score greater than or equal to the score of the result being printed, and is derived by analysis of the total score distribution.

SUMMARIES

Result No.	Score	% Query Match	Length	DB	ID	Description
1	10	100.0	3	2	AAW56247	Aaw56247 Anti-infl
2	10	100.0	3	5	ABG77780	Abg77780 Targettin
3	10	100.0	4	1	AAP91627	Aap91627 Motif use
4	10	100.0	4	2	AAR41636	Aar41636 Internali
5	10	100.0	4	2	AAR54662	Aar54662 Native se
6	10	100.0	4	2	AAR93665	Aar93665 HIV princ
7	10	100.0	4	2	AAW86792	Aaw86792 Tetrapept
8	10	100.0	4	2	AAW77469	Aaw77469 Tetrapept
9	10	100.0	4	2	AAW37726	Aaw37726 Cytochrom
10	10	100.0	4	2	AAW46523	Aaw46523 Peptide c
11	10	100.0	4	2	AAW87483	Aaw87483 HIV-1 MO/
12	10	100.0	4	2	AAW87477	Aaw87477 HIV-1 mut
13	10	100.0	4	2	AAY21283	Aay21283 Human sem
14	10	100.0	4	2	AAW55768	Aaw55768 Immunisat
15	10	100.0	4	2	AAW47950	Aaw47950 AE101 ana
16	10	100.0	4	3	AAY85398	Aay85398 IL-2 deri
17	10	100.0	4	3	AAY85365	Aay85365 IL-2 deri
18	10	100.0	4	3	AAY85397	Aay85397 IL-2 deri
19	10	100.0	4	3	AAY85401	Aay85401 IL-2 deri
20	10	100.0	4	3	AAB14508	Aab14508 Sterile a
21	10	100.0	4	3	AAY99633	Aay99633 Insectici
22	10	100.0	4	4	AAB86380	Aab86380 Cathepsin
23	10	100.0	4	4	AAB73471	Aab73471 Mammalian
24	10	100.0	4	5	AAU99275	Aau99275 Synthetic
25	10	100.0	4	5	AAU73263	Aau73263 Human pro
26	10	100.0	4	5	ABB84331	Abb84331 Human MBP
27	10	100.0	4	6	ABJ26706	Abj26706 Seed deve
28	10	100.0	4	6	ABB99613	Abb99613 Peptide d
29	10	100.0	4	7	ADW36990	Adw36990 HLA bindi
30	10	100.0	4	8	ADM12048	Adm12048 Modified
31	10	100.0	4	8	ADM12823	Adm12823 Ii key/MH
32	10	100.0	4	8	ADM79799	Adm79799 Pig angio
33	10	100.0	4	8	ADO38266	Ado38266 Modified
34	10	100.0	4	8	ADO38670	Ado38670 Melanocyt
35	10	100.0	4	8	ADO39060	Ado39060 Ebola vir
36	10	100.0	4	9	ADY86552	Ady86552 Human VEG
37	10	100.0	4	9	ADY86539	Ady86539 Murine VE
38	10	100.0	4	9	AEB09478	Aeb09478 Angiotens
39	10	100.0	5	2	AAR09365	Aar09365 Sequence
40	10	100.0	5	2	AAR09371	Aar09371 Sequence
41	10	100.0	5	2	AAR25238	Aar25238 PTHrP (10
42	10	100.0	5	2	AAR71657	Aar71657 Natriuret
43	10	100.0	5	2	AAR54661	Aar54661 Native se
44	10	100.0	5	2	AAW02183	Aaw02183 Residues
45	10	100.0	5	2	AAW04336	Aaw04336 ATP dipho

ALIGNMENTS

RESULT 1
AAW56247
ID AAW56247 standard; peptide; 3 AA.
XX
AC AAW56247;
XX
DT 20-JUL-1998 (first entry)
XX
DE Anti-inflammatory tripeptide.
XX
KW Anti-inflammatory; macrophage inhibitory activity; fibronectin;
KW T-cell inhibitory activity; adherence; extracellular matrix;
KW up-regulation; fas receptor expression; inflammation.
XX
OS Synthetic.
XX
PN WO9809985-A2.
XX
PD 12-MAR-1998.
XX
PF 03-SEP-1997; 97WO-IL000295.
XX
PR 03-SEP-1996; 96US-0025376P.
PR 20-NOV-1996; 96US-00753141.
PR 28-MAY-1997; 97US-00864301.
XX
PA (YEDA) YEDA RES & DEV CO LTD.
XX
PI Eisenbachschwartz M, Beserman P, Hirschberg DL;
XX
DR WPI; 1998-193550/17.
XX
PT Anti-inflammatory peptides and derivatives - used for treating, e.g.
PT arthritis, ulcerative colitis, auto-immune disease, allergy asthma,
PT shock, HIV infection, transplant rejection or Alzheimer's disease.
XX
PS Claim 7; Page 35; 42pp; English.
XX
CC AAW56171-248 represent anti-inflammatory tripeptides of the invention.
CC They are derived from the formulae: Xaa-Glu-Arg, Arg-Glu-Xaa, Xaa-Arg-
CC Glu, or Glu-arg-Xaa, where Xaa = any amino acid residue. Cyclic
CC derivatives of the peptides also function as anti-inflammatory agents.
CC The peptides can be covalently linked to one another either directly or
CC through a spacer. The peptides and their derivatives have macrophage
CC inhibitory and T-cell inhibitory activity and thus, anti-inflammatory
CC activity. The peptides and compositions have anti-immune activity, i.e.
CC inhibitory effects against a cellular and humoral immune response,

CC including a response not associated with inflammation. The peptides also
CC inhibit the ability of macrophages and T-cells to adhere to extracellular
CC matrix components and fibronectin, as well as up-regulated fas receptor
CC expression in T-cells. They can be used to inhibit unwanted immune
CC reaction and inflammation
XX
SQ Sequence 3 AA;

Query Match 100.0%; Score 10; DB 2; Length 3;
Best Local Similarity 100.0%; Pred. No. 2e+06;
Matches 2; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

Qy 1 RM 2
||
Db 2 RM 3

RESULT 2
ABG77780
ID ABG77780 standard; peptide; 3 AA.
XX
AC ABG77780;
XX
DT 05-NOV-2002 (first entry)
XX
DE Targetting peptide selective for human organ, tissue or cell type #313.
XX
KW Human; cytostatic; antiinflammatory; antidiabetic; cardiovascular;
KW immunomodulator; antibacterial; antiviral; gene therapy; cancer;
KW arthritis; diabetes; inflammatory disease; atherosclerosis;
KW autoimmune disease; bacterial infection; viral infection;
KW cardiovascular disease; degenerative disease.
XX
OS Homo sapiens.
XX
PN WO200220723-A2.
XX
PD 14-MAR-2002.
XX
PF 07-SEP-2001; 2001WO-US028044.
XX
PR 08-SEP-2000; 2000US-0231266P.
PR 17-JAN-2001; 2001US-00765101.
XX
PA (TEXA) UNIV TEXAS SYSTEM.
XX
PI Arap W, Pasqualini R;
XX
DR WPI; 2002-599247/64.
XX
PT New targeting peptides identified by phage display, useful for treating a
PT disease state, e.g. cancer, diabetes, inflammatory disease,
PT atherosclerosis, autoimmune disease, bacterial or viral infection or
PT cardiovascular disease.
XX
PS Claim 16; Page 78; 269pp; English.
XX
CC The invention describes an isolated peptide of 100 amino acids or less in
CC size. The peptide is useful for treating a disease state, e.g. cancer,
CC arthritis, diabetes, inflammatory disease, atherosclerosis, autoimmune
CC disease, bacterial infection, viral infection, cardiovascular disease or
CC degenerative disease. This sequence represents a human targeting peptide
CC selective for human organs, tissues or cell types
XX
SQ Sequence 3 AA;

Query Match 100.0%; Score 10; DB 5; Length 3;
Best Local Similarity 100.0%; Pred. No. 2e+06;
Matches 2; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

Qy 1 RM 2
||
Db 2 RM 3

RESULT 3
AAP91627
ID AAP91627 standard; protein; 4 AA.
XX
AC AAP91627;
XX
DT 25-MAR-2003 (revised)
DT 09-JUL-1990 (first entry)
XX
DE Motif useful in tolerization alone or in association with epitopes to a
DE proteolipid protein.
XX
KW Autoantigen; MBP; myelin basic protein; transplantation antigen;
KW myasthenia gravis; myasthenics; Transplantation antigen.
XX
OS Synthetic.
XX
PN EP304279-A.
XX
PD 22-FEB-1989.
XX
PF 17-AUG-1988; 88EP-00307608.
XX
PR 17-AUG-1987; 87US-00086694.
XX
PA (STRD) UNIV LELAND STANFORD JUNIOR.
XX
PI Steinman L, Zamvil S;
XX
DR WPI; 1989-055696/08.
XX
PT Oligopeptide and polypeptide compsns. - based on the amino acid sequence
PT of an immunogen and used for modulating the immune system.
XX
PS Disclosure; Page; 7pp; English.
XX
CC Sequences will normally be part of 9-15 amino acid sequence, excluded as
CC motifs for immunisation but useful in tolerisation. . (Updated on 25-MAR-
CC 2003 to correct PR field.) (Updated on 25-MAR-2003 to correct PA field.)
CC (Updated on 25-MAR-2003 to correct PI field.)
XX
SQ Sequence 4 AA;

Query Match 100.0%; Score 10; DB 1; Length 4;
Best Local Similarity 100.0%; Pred. No. 2e+06;
Matches 2; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

Qy 1 RM 2
||
Db 1 RM 2

RESULT 4
AAR41636
ID AAR41636 standard; peptide; 4 AA.
XX
AC AAR41636;
XX
DT 25-MAR-2003 (revised)
DT 10-MAR-1994 (first entry)
XX
DE Internalisation signal #1.
XX
KW Internalisation signal; core; modulation; receptor; transport; ligand;
KW cytoplasmic tail; endocytosis.
XX
OS Synthetic.
XX
PN WO9318185-A1.
XX
PD 16-SEP-1993.

XX
PF 01-MAR-1993; 93WO-US001669.
XX
PR 03-MAR-1992; 92US-00844852.
XX
PA (SALK) SALK INST BIOLOGICAL STUDIES.
PA (SCRI) SCRIPPS RES INST.
XX
PI Trowbridge IS, Collawn JF, Tainer JA, Kuhn LA;
XX
DR WPI; 1993-303496/38.
XX
PT Modulating receptor mediated transport of ligand into cell - by
PT introducing heterologous internalisation signal into cell.
XX
PS Claim 16; Page 49; 60pp; English.
XX
CC The sequences given in AAR41636-57 represent the cores of internalisation
CC signals which were used in the method of the invention for modulating
CC receptor mediated transport of a ligand into a cell. These sequences are
CC derived from the cytoplasmic tails of surface receptors. These amino acid
CC internalisation signals have a tight turn structure. The introduction of
CC one of these sequences into a receptor within a cell, modulates the
CC transport of ligand into a cell having a surface receptor reactive with
CC that ligand. This modulation can cause an increase or a decrease in
CC endocytosis, depending on the choice of internalisation signal. (Updated
CC on 25-MAR-2003 to correct PN field.)
XX
SQ Sequence 4 AA;

Query Match 100.0%; Score 10; DB 2; Length 4;
Best Local Similarity 100.0%; Pred. No. 2e+06;
Matches 2; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

Qy 1 RM 2
||
Db 3 RM 4

RESULT 5
AAR54662
ID AAR54662 standard; peptide; 4 AA.
XX
AC AAR54662;
XX
DT 25-MAR-2003 (revised)
DT 29-NOV-1994 (first entry)
XX
DE Native secreted amyloid precursor protein (APP) antagonistic peptide.
XX
KW Amyloid precursor protein- antagonist; Alzheimer's disease;
KW neuron growth.
XX
OS Synthetic.
XX
PN WO9409808-A1.
XX
PD 11-MAY-1994.
XX
PF 23-OCT-1992; 92WO-US009070.
XX
PR 23-OCT-1992; 92WO-US009070.
XX
PA (REGC) UNIV CALIFORNIA.
XX
PI Saitoh T;
XX
DR WPI; 1994-167118/20.
XX
PT Peptide(s) and analogues based on amyloid precursor protein - used for
PT promoting neuronal growth in conditions involving damage to neurons or in
PT treating Alzheimer's Disease etc.
XX
PS Claim 52; Page 11; 116pp; English.
XX
CC This sequence corresponds to AA 330-333 of amyloid precursor protein.
CC This peptide, which partially overlaps the C-terminal side of thet active
CC sequence RERMS (AAR54661), acts to antagonize the activity of secreted
CC APP (695 AA). (Updated on 25-MAR-2003 to correct PN field.)
XX
SQ Sequence 4 AA;

Query Match 100.0%; Score 10; DB 2; Length 4;
Best Local Similarity 100.0%; Pred. No. 2e+06;
Matches 2; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

Qy 1 RM 2
||
Db 1 RM 2

RESULT 6
AAR93665
ID AAR93665 standard; peptide; 4 AA.
XX
AC AAR93665;
XX
DT 27-APR-1996 (first entry)
XX
DE HIV principal neutralisation epitope binding to 559 antibody.
XX
KW SPNE; selected principal neutralisation epitope; vaccine; HIV;
KW outer membrane proteosome; Neisseria; OMPC; AIDS; 559 antibody.
XX
OS Synthetic.
XX
PN GB2282380-A.
XX
PD 05-APR-1995.
XX
PF 23-SEP-1994; 94GB-00019256.
XX
PR 30-SEP-1993; 93US-00129720.
XX
PA (MERI) MERCK & CO INC.
XX
PI Arnold BA, Conley AJ, Keller PM;
XX
DR WPI; 1995-125267/17.
XX
PT New antigenic conjugate useful as vaccine for AIDS - comprising HIV
PT principal neutralisation epitope covalently linked to outer membrane
PT proteosome of Neisseria.
XX
PS Claim 14; Page 9; 66pp; English.
XX
CC An antigenic conjugate, useful as a vaccine for AIDS, has the formula
CC (SPNE)n-(OMPC), where SPNE is a selected principal neutralisation epitope
CC of HIV, which is one of 16 specified polypeptides (including the present
CC sequence) or their fragments containing at least 5 amino acids; OMPC is
CC purified outer membrane proteosome of Neisseria (pref. N. meningitidis);
CC and n is 1-200, indicating the number of SPNE moieties covalently linked
CC to the OMPC. The conjugates may be substituted by anions, and conjugation
CC may be via a bigeneric spacer. The SPNE polypeptides bind an HIV broadly
CC neutralising monoclonal antibody (559 antibody) that binds gp120 and also
CC inhibits the binding of HIV to CD4. They were originally identified in
CC the screening of phage epitope libraries having randomly generated
CC epitope polypeptides accessible to the antibody. The libraries were ALPHA
CC and EPSILON described in AAR83295 and AAR83299 respectively.The sequences
CC of these polypeptides were deduced from their corresponding DNA sequence,
CC determined by PCR
XX
SQ Sequence 4 AA;

Query Match 100.0%; Score 10; DB 2; Length 4;
Best Local Similarity 100.0%; Pred. No. 2e+06;

```
  Matches    2;  Conservative    0;  Mismatches    0;  Indels    0;  Gaps    0;

Qy          1 RM 2
              ||
Db          2 RM 3


RESULT 7
AAW86792
ID   AAW86792 standard; peptide; 4 AA.
XX
AC   AAW86792;
XX
DT   06-JUL-1999  (first entry)
XX
DE   Tetrapeptide useful as zinc endopeptidase 24-15 inhibitor.
XX
KW   Zinc endopeptidase; EC.3.4.24-15; selective inhibitor; analgesic; pain;
KW   hypothermia; arterial hypertension; cancer; Alzheimer's disease;
KW   phosphinic acid; pseudopeptide linkage.
XX
OS   Synthetic.
XX
FH   Key             Location/Qualifiers
FT   Modified-site   1. .2
FT                   /note= "Z-Phe-psi[PO2CH2]-Ala, where Z is benzyloxy-
FT                   carbonyl and -psi[PO2CH2]- indicates replacement of the
FT                   peptide linkage -CONH- between Phe and Ala by the group -
FT                   PO2CH2-"
XX
PN   EP725075-A1.
XX
PD   07-AUG-1996.
XX
PF   02-FEB-1996;   96EP-00400229.
XX
PR   06-FEB-1995;   95FR-00001328.
XX
PA   (COMS ) COMMISSARIAT ENERGIE ATOMIQUE.
XX
PI   Dive V,  Jiracek J,  Yiotakis A;
XX
DR   WPI; 1996-356059/36.
XX
PT   New peptide derivs. contg. phosphonic acid gp. replacing an amide bond -
PT   are highly specific inhibitors of endo:peptidase 24-15, for treating
PT   hypothermia, hypertension, cancer, Alzheimer's disease etc.
XX
PS   Disclosure; Page 12; 18pp; French.
XX
CC   The sequence is a specific example of new peptide derivatives containing
CC   the sequence -Phe-psi[PO2CH2]-X-Y-Z'- in which Y = Arg or Lys; X and Z'=
CC   natural or pseudo-amino acids (preferably X is Gly, Ala or Leu and Z' is
CC   Met, Nle, Ala or Phe); and -psi[PO2CH2]- indicates replacement of the
CC   peptide linkage -CONH- between Phe and X by the group -PO2CH2-. These
CC   peptides are inhibitors of the zinc-dependent endopeptidase EC.3.4.24-15
CC   and so prevent degradation of e.g. somatostatin, bradykinin, angiotensin,
CC   neurotensin, substance P, dynorphin etc. and may prevent maturation of
CC   ras oncoprotein. They are useful in treatment of pain, hypothermia,
CC   arterial hypertension, cancer and Alzheimer's disease. They are very
CC   selective for 24-15 with no significant action on other zinc
CC   endopeptidases such as 24-16, and are more stable, chemically, than
CC   phosphonamide peptide derivatives
XX
SQ   Sequence 4 AA;

  Query Match             100.0%;  Score 10;  DB 2;  Length 4;
  Best Local Similarity   100.0%;  Pred. No. 2e+06;
  Matches    2;  Conservative    0;  Mismatches    0;  Indels    0;  Gaps    0;

Qy          1 RM 2
              ||
Db          3 RM 4


RESULT 8
AAW77469
ID   AAW77469 standard; peptide; 4 AA.
XX
AC   AAW77469;
XX
DT   06-JUL-1999  (first entry)
XX
DE   Tetrapeptide useful as zinc endopeptidase 24-15 inhibitor.
XX
KW   Zinc endopeptidase; EC.3.4.24-15; selective inhibitor; analgesic; pain;
KW   hypothermia; arterial hypertension; cancer; Alzheimer's disease;
KW   phosphinic acid; pseudopeptide linkage.
XX
OS   Synthetic.
XX
FH   Key             Location/Qualifiers
FT   Modified-site   1. .2
FT                   /note= "Z-Phe-psi[PO2CH2]-Gly, where Z is benzyloxy-
FT                   carbonyl and -psi[PO2CH2]- indicates replacement of the
FT                   peptide linkage -CONH- between Phe and Gly by the group -
FT                   PO2CH2-"
XX
PN   EP725075-A1.
XX
PD   07-AUG-1996.
XX
PF   02-FEB-1996;   96EP-00400229.
XX
PR   06-FEB-1995;   95FR-00001328.
XX
PA   (COMS ) COMMISSARIAT ENERGIE ATOMIQUE.
XX
PI   Dive V,  Jiracek J,  Yiotakis A;
XX
DR   WPI; 1996-356059/36.
XX
PT   New peptide derivs. contg. phosphonic acid gp. replacing an amide bond -
PT   are highly specific inhibitors of endo:peptidase 24-15, for treating
PT   hypothermia, hypertension, cancer, Alzheimer's disease etc.
XX
PS   Claim 7; Page 13; 18pp; French.
XX
CC   The sequence is a specific example of new peptide derivatives containing
CC   the sequence -Phe-psi[PO2CH2]-X-Y-Z'- in which Y = Arg or Lys; X and Z'=
CC   natural or pseudo-amino acids (preferably X is Gly, Ala or Leu and Z' is
CC   Met, Nle, Ala or Phe); and -psi[PO2CH2]- indicates replacement of the
CC   peptide linkage -CONH- between Phe and X by the group -PO2CH2-. These
CC   peptides are inhibitors of the zinc-dependent endopeptidase EC.3.4.24-15
CC   and so prevent degradation of e.g. somatostatin, bradykinin, angiotensin,
CC   neurotensin, substance P, dynorphin etc. and may prevent maturation of
CC   ras oncoprotein. They are useful in treatment of pain, hypothermia,
CC   arterial hypertension, cancer and Alzheimer's disease. They are very
CC   selective for 24-15 with no significant action on other zinc
CC   endopeptidases such as 24-16, and are more stable, chemically, than
CC   phosphonamide peptide derivatives
XX
SQ   Sequence 4 AA;

  Query Match             100.0%;  Score 10;  DB 2;  Length 4;
  Best Local Similarity   100.0%;  Pred. No. 2e+06;
  Matches    2;  Conservative    0;  Mismatches    0;  Indels    0;  Gaps    0;

Qy          1 RM 2
              ||
Db          3 RM 4


RESULT 9
AAW37726
ID   AAW37726 standard; peptide; 4 AA.
```

XX
AC AAW37726;
XX
DT 07-JUL-1998 (first entry)
XX
DE Cytochrome P450 degenerate primer 2.
XX
KW Degenerate peptide; PCR; amplification; cytochrome P450 gene;
KW oxidative metabolism; P450RAI; retinoic acid; RA; promoter.
XX
OS Synthetic.
XX
PN WO9749832-A2.
XX
PD 31-DEC-1997.
XX
PF 23-JUN-1997; 97WO-CA000488.
XX
PR 21-JUN-1996; 96US-00667546.
PR 01-OCT-1996; 96US-00724466.
XX
PA (TOOH) UNIV QUEENS KINGSTON.
XX
PI Petkovich PM;
XX
DR WPI; 1998-077193/07.
XX
PT Identifying DNA encoding inducible or suppressible cytochrome P450 - by
PT screening for drugs which reduce the catabolism of retinoic acid, useful
PT in cancer chemotherapy and the treatment of acne and psoriasis.
XX
PS Claim 17; Page 61; 113pp; English.
XX
CC This is a degenerate peptide from which the PCR primer used in the
CC amplification process is derived. The PCR involves the inducible
CC cytochrome P450 gene which produces enzymes involved in the oxidative
CC metabolism of endogenous and exogenous compounds. The cytochrome P450
CC nucleotide sequence can be used to induce or suppress the expression of
CC its protein. A form of cytochrome P450 is P450RAI which specifically
CC metabolises a derivative of retinoic acid (RA). It is highly induced by
CC RA in cell lines and tissues. This allows for the development of a drug
CC screen using promoters and nucleotide sequences to identify drugs which
CC are useful for reducing the catabolism of RA
XX
SQ Sequence 4 AA;

Query Match 100.0%; Score 10; DB 2; Length 4;
Best Local Similarity 100.0%; Pred. No. 2e+06;
Matches 2; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

Qy 1 RM 2
||
Db 3 RM 4

RESULT 10
AAW46523
ID AAW46523 standard; peptide; 4 AA.
XX
AC AAW46523;
XX
DT 20-MAY-1998 (first entry)
XX
DE Peptide containing a protease binding site.
XX
KW Protease binding site; protease; protease indicator; fluorescent signal;
KW detection; protease activity.
XX
OS Synthetic.
XX
PN US5714342-A.
XX
PD 03-FEB-1998.
XX
PF 27-OCT-1995; 95US-00549008.
XX
PR 28-OCT-1994; 94US-00331383.
XX
PA (ONCO-) ONCOIMMUNIN INC.
XX
PI Packard BS, Komoriya A;
XX
DR WPI; 1998-158345/14.
XX
PT Fluorogenic substrates for protease determination - having two closely
PT spaced fluorophores flanking protease binding site.
XX
PS Disclosure; Col 4; 39pp; English.
XX
CC Peptides AAW46520-53 contain protease binding sites. They are used to
CC produce novel reagents whose fluorescence increases in the presence of
CC particular proteases. These fluorogenic protease indicators (substrates)
CC provide a high intensity fluorescent signal at a visible wavelength when
CC they are digested by a protease. The fluorogenic indicators have the
CC general formula: F1--C1--P--C2--F2 | | (S1)n (S2)k where: P is a peptide
CC containing a protease binding site, e.g. AAW46520-53. F1 and F2 are
CC fluorophores. S1 and S2 are peptide spacers e.g. AAW46554-58. n, k = 0 or
CC 1. C1 and C2 are conformation-determining regions that introduce a bend
CC into the composition which positions the fluorophores adjacent to each
CC other with a separation of less than 100 Angstrom. When n is 1, S1 is
CC joined to the terminal alpha -amino group of C1 by a peptide bond, and
CC when k is 1, S2 is joined to the terminal carboxy group of C2 by a
CC peptide bond. The protease indicators are used for detecting protease
CC activity in a biological sample. The sample is contacted with the
CC indicator and any change in fluorescence is detected, an increase in
CC fluorescence indicating protease activity
XX
SQ Sequence 4 AA;

Query Match 100.0%; Score 10; DB 2; Length 4;
Best Local Similarity 100.0%; Pred. No. 2e+06;
Matches 2; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

Qy 1 RM 2
||
Db 1 RM 2

RESULT 11
AAW87483
ID AAW87483 standard; peptide; 4 AA.
XX
AC AAW87483;
XX
DT 17-OCT-2003 (revised)
DT 11-FEB-1999 (first entry)
XX
DE HIV-1 MO/LAI env gp120 amino acid fragment.
XX
KW Human immunodeficiency virus-1; HIV-1; protease; provirus; L-2 cell;
KW mutation; Nef protein; immunogen; apoptosis; treatment; hybridisation;
KW cell lysis; chemotherapeutic; toxin; HIV infection; MO/LAI.
XX
OS Human immunodeficiency virus 1.
XX
FH Key Location/Qualifiers
FT Misc-difference 2
FT /note= "encoded by GAA"
XX
PN WO9844945-A1.
XX
PD 15-OCT-1998.
XX
PF 03-APR-1998; 98WO-US006690.
XX
PR 04-APR-1997; 97US-0043047P.

```
XX
PA   (IMMU-) IMMUNE RESPONSE CORP.
XX
PI   Luftig RB;
XX
DR   WPI; 1998-568304/48.
DR   N-PSDB; AAV71866.
XX
PT   New nucleic acid encoding mutant or truncated forms of human immune
PT   deficiency virus proteins - used to generate non-infectious particles
PT   useful as therapeutic or prophylactic immunogens, also for diagnosis.
XX
PS   Example 1; Fig 1; 66pp; English.
XX
CC   AAW87481 to AAW87486 represent human immunodeficiency virus-1 (HIV-1)
CC   sequences in MO/LAI cells that were used for comparing the mutant amino
CC   acid sequences of the defective HIV provirus in L-2 cells. AAV71858 to
CC   AAV71863 represent the portions of pol protease (prot.), vpr, env (gp120
CC   and gp41), and nef gene regions that were mutated as compared to wild-
CC   type HIV-1 in LAI or MO/LAI cells. The invention relates to mutated DNA,
CC   proteins or fragments of defective HIV-1 provirus in L-2 cell line. The
CC   mutations result in non-infectious HIV-1 particles. Inactive, protease-
CC   deficient HIV-1 particles containing at least one of Nef truncated
CC   protein fragments are used as immunogens, particularly for reducing or
CC   preventing apoptosis in HIV-1 sero-negative or -positive subjects,
CC   specifically those with HIV-1 infection, both for prevention and
CC   treatment. Fragments of nef gene and fragments encoding specific mutant
CC   Nef proteins are also useful in hybridisation tests for diagnostic
CC   detection of mutated genes in (lysed) cells or body fluids, while the
CC   corresponding mutant proteins are detected in immunoassays using
CC   antibodies. The protease-deficient HIV-1 particles and antibodies,
CC   optionally attached to a radioisotope, chemotherapeutic agent or toxin,
CC   can be used to reduce the severity of HIV infections. (Updated on 17-OCT-
CC   2003 to standardise OS field)
XX
SQ   Sequence 4 AA;

  Query Match             100.0%;  Score 10;  DB 2;  Length 4;
  Best Local Similarity   100.0%;  Pred. No. 2e+06;
  Matches    2;  Conservative    0;  Mismatches    0;  Indels    0;  Gaps    0;

Qy          1 RM 2
              ||
Db          2 RM 3


RESULT 12
AAW87477
ID   AAW87477 standard; peptide; 4 AA.
XX
AC   AAW87477;
XX
DT   17-OCT-2003  (revised)
DT   11-FEB-1999  (first entry)
XX
DE   HIV-1 mutant env gp120 amino acid fragment in L-2 cell.
XX
KW   Human immunodeficiency virus-1; HIV-1; protease; provirus; L-2 cell;
KW   mutation; Nef protein; immunogen; apoptosis; treatment; hybridisation;
KW   cell lysis; chemotherapeutic; toxin; HIV infection.
XX
OS   Human immunodeficiency virus 1.
XX
FH   Key             Location/Qualifiers
FT   Misc-difference 2
FT                   /note= "encoded by GAA"
XX
PN   WO9844945-A1.
XX
PD   15-OCT-1998.
XX
PF   03-APR-1998;   98WO-US006690.
XX
PR   04-APR-1997;   97US-0043047P.
XX
PA   (IMMU-) IMMUNE RESPONSE CORP.
XX
PI   Luftig RB;
XX
DR   WPI; 1998-568304/48.
DR   N-PSDB; AAV71860.
XX
PT   New nucleic acid encoding mutant or truncated forms of human immune
PT   deficiency virus proteins - used to generate non-infectious particles
PT   useful as therapeutic or prophylactic immunogens, also for diagnosis.
XX
PS   Example 1; Fig 1; 66pp; English.
XX
CC   Sequences AAW87475 to AAW87463 represent mutant amino acid sequences of
CC   the defective human immunodeficiency virus-1 (HIV-1) provirus in L-2
CC   cells. They represent the portions of pol protease (prot.), vpr, env
CC   (gp120 and gp41), and nef gene regions that were mutated as compared to
CC   wild-type HIV-1 in LAI or MO/LAI cells. The invention relates to mutated
CC   DNA, proteins or fragments of defective HIV-1 provirus in L-2 cell line.
CC   The mutations result in non-infectious HIV-1 particles. Inactive,
CC   protease-deficient HIV-1 particles containing at least one of Nef
CC   truncated protein fragments are used as immunogens, particularly for
CC   reducing or preventing apoptosis in HIV-1 sero-negative or -positive
CC   subjects, specifically those with HIV-1 infection, both for prevention
CC   and treatment. Fragments of nef gene and fragments encoding specific
CC   mutant Nef proteins are also useful in hybridisation tests for diagnostic
CC   detection of mutated genes in (lysed) cells or body fluids, while the
CC   corresponding mutant proteins are detected in immunoassays using
CC   antibodies. The protease-deficient HIV-1 particles and antibodies,
CC   optionally attached to a radioisotope, chemotherapeutic agent or toxin,
CC   can be used to reduce the severity of HIV infections. (Updated on 17-OCT-
CC   2003 to standardise OS field)
XX
SQ   Sequence 4 AA;

  Query Match             100.0%;  Score 10;  DB 2;  Length 4;
  Best Local Similarity   100.0%;  Pred. No. 2e+06;
  Matches    2;  Conservative    0;  Mismatches    0;  Indels    0;  Gaps    0;

Qy          1 RM 2
              ||
Db          2 RM 3


RESULT 13
AAY21283
ID   AAY21283 standard; protein; 4 AA.
XX
AC   AAY21283;
XX
DT   22-JUL-1999  (first entry)
XX
DE   Human semaphorin III mutant protein fragment 16.
XX
KW   Human; beta-amyloid precursor protein; beta-APP; diagnosis; cancer;
KW   frameshift mutation; age-related disease; neurodegenerative disorder;
KW   Alzheimer's disease; Down's syndrome; myotonic dystrophy; neuronal;
KW   Huntington's disease; multiple sclerosis; alcoholic liver disease;
KW   diabetes mellitus type II; microtubule associated protein; Tau; Big Tau;
KW   ubiquitin B; apolipoprotein E; MAP2; neurofilament-L; neurofilament-M;
KW   neurofilament-F; presenilin I; presenilin II; cellular tumour antigen;
KW   glial fibrillary acidic protein; GFAP; p53; semaphorin III; HUPF-1;
KW   bcl-2; B-cell leukemia/lymphoma 2 proto-oncogene; HMGP-C; NSP-A;
KW   high mobility group protein-C; neuroendocrine specific protein A.
XX
OS   Synthetic.
OS   Homo sapiens.
XX
PN   WO9845322-A2.
XX
PD   15-OCT-1998.
```

XX
PF 02-APR-1998; 98WO-IB000705.
XX
PR 10-APR-1997; 97US-0043163P.
XX
PA (ROYA-) ROYAL NETHERLANDS ACAD ARTS & SCI.
PA (UYRO-) UNIV ROTTERDAM ERASMUS.
PA (UYUT-) RIJKSUNIV UTRECHT.
XX
PI Van Leeuwen FW, Grosveld FG, Burbach JPH;
XX
DR WPI; 1998-609901/51.
DR N-PSDB; AAX75767.
XX
PT Diagnosing disease by detecting frameshift mutations in RNA or
PT corresponding protein mutations - used to diagnose cancer and
PT neurological diseases, particularly Alzheimer's disease, and also for
PT treatment and prevention with specific ribozymes or wild-type RNA.
XX
PS Disclosure; Fig 16; 258pp; English.
XX
CC This invention describes a novel method for the diagnosis of a disease
CC caused by, or associated with, an RNA molecule that has a frameshift
CC mutation. The method is used to diagnose age-related diseases, especially
CC cancer and a wide range of neurodegenerative disorders (e.g. Alzheimer's
CC disease, Down's syndrome, myotonic dystrophy, Huntington's disease,
CC multiple sclerosis, alcoholic liver disease, diabetes mellitus type II
CC and many others listed) or susceptibility to these disorders. The method
CC allows a definitive diagnosis of Alzheimer's disease in living patients,
CC at an early stage. It is based on the observation that disease may be
CC caused by mutations in RNA rather than DNA. The invention describes the
CC used of neuronal system RNA molecules, specifically proteins including
CC beta-amyloid precursor protein (beta-APP), the microtubule associated
CC proteins Tau and Big Tau, ubiquitin B, apolipoprotein E, microtubule
CC associated protein 2 (MAP2), neurofilament-L, neurofilament-M,
CC neurofilament-F, presenilin I, presenilin II, glial fibrillary acidic
CC protein (GFAP), the cellular tumour antigen p53, B-cell leukemia/lymphoma
CC 2 (bcl-2) proto-oncogene, semaphorin III, HUPF-1, high mobility group
CC protein-C (HMGP-C) and neuroendocrine specific protein A
XX
SQ Sequence 4 AA;

Query Match 100.0%; Score 10; DB 2; Length 4;
Best Local Similarity 100.0%; Pred. No. 2e+06;
Matches 2; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

Qy 1 RM 2
||
Db 3 RM 4

RESULT 14
AAW55768
ID AAW55768 standard; peptide; 4 AA.
XX
AC AAW55768;
XX
DT 25-MAR-2003 (revised)
DT 08-JUL-1998 (first entry)
XX
DE Immunisation motif associated with PLP 7.
XX
KW Myelin basic protein; immunity; immune response; neurological; T-cell;
KW human; immunogen; B-cell; transplantation antigen; immunomodulator.
XX
OS Unidentified.
XX
PN EP805162-A1.
XX
PD 05-NOV-1997.
XX
PF 17-AUG-1988; 97EP-00106788.
XX
PR 17-AUG-1987; 87US-00086694.
PR 17-AUG-1988; 88EP-00307608.
XX
PA (STRD) UNIV LELAND STANFORD JUNIOR.
XX
PI Steinman L, Zamvil S;
XX
DR WPI; 1998-034664/04.
XX
PT Polypeptide comprising human myelin basic protein fragment - useful as
PT immuno modulator.
XX
PS Disclosure; Page 8; 8pp; English.
XX
CC The present sequence represents an immunisation motif normally excluded,
CC but which may be used with advantage for tolerisation by itself or in
CC conjunction with other epitope sequences from the present invention. The
CC present invention describes a polypeptide comprising a human myelin basic
CC protein (hMBP) fragment including P89-101 of hMBP, excluding native hMBP.
CC The term P89-101 is not defined but may be intended to mean amino acids
CC 89-101 of hMBP. The polypeptide can be used for tolerising a mammalian
CC host immune system comprising B and T cells to an immunogen of interest,
CC wherein said immunogen is restricted by a transplantation antigen of said
CC host. (Updated on 25-MAR-2003 to correct PF field.) (Updated on 25-MAR-
CC 2003 to correct PR field.)
XX
SQ Sequence 4 AA;

Query Match 100.0%; Score 10; DB 2; Length 4;
Best Local Similarity 100.0%; Pred. No. 2e+06;
Matches 2; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

Qy 1 RM 2
||
Db 1 RM 2

RESULT 15
AAW47950
ID AAW47950 standard; peptide; 4 AA.
XX
AC AAW47950;
XX
DT 12-JUN-1998 (first entry)
XX
DE AE101 analogue effector compound SEQ ID NO:26.
XX
KW Mammalian Ii key peptide; mammalian invariant chain protein; allergy;
KW immune response; MHC class II; antigenic; autoimmune disease.
XX
OS Synthetic.
XX
FH Key Location/Qualifiers
FT Modified-site 1
FT /note= "Acylated"
FT Modified-site 4
FT /note= "Amidated"
XX
PN WO9749430-A1.
XX
PD 31-DEC-1997.
XX
PF 09-JUN-1997; 97WO-US009993.
XX
PR 26-JUN-1996; 96US-00670605.
XX
PA (ANTI-) ANTIGEN EXPRESS INC.
XX
PI Humphreys RE, Adams S, Xu M;
XX
DR WPI; 1998-076917/07.
XX
PT New mammalian invariant chain protein (Ii) key peptide(s) - used for

PT modulation of immune response, e.g. for treating malignant, allergic or
PT autoimmune disease or allograft rejection.
XX
PS Example 1; Page 27; 149pp; English.
XX
CC The present sequence represents an AE101 analogue effector compound used
CC in the present invention. The present invention describes a mammalian
CC invariant chain protein (Ii) key peptide of sequence LRMKLPKPPKPVSKMR and
CC modifications with the exclusion of peptide YRMKLPKPPKPVSKMR. MHC class
CC II molecules are synthesised in the endoplasmic reticulum with their
CC antigenic peptide sites blocked by the invariant chain protein (Ii). The
CC products and method can be used for the modulation of an immune response
CC for therapeutic or diagnostic purposes. The enhancement of immunity can
CC be used in the treatment of e.g. malignant or allergic disease. The
CC immunosuppression can be used for the treatment of autoimmune disease,
CC e.g. rheumatoid arthritis, multiple sclerosis, diabetes mellitus, lupus
CC erythematosus, and psoriasis or allograft rejection
XX
SQ Sequence 4 AA;

Query Match 100.0%; Score 10; DB 2; Length 4;
Best Local Similarity 100.0%; Pred. No. 2e+06;
Matches 2; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

Qy 1 RM 2
||
Db 2 RM 3

Search completed: February 28, 2006, 08:45:08
Job time : 2.68505 secs

GenCore version 5.1.7

OM protein - protein search, using sw model

Run on: February 28, 2006, 08:53:26 ; Search time 0.172097 Seconds
(without alignments)
960.804 Million cell updates/sec

Title: US-10-717-665A-44_COPY_448_449
Perfect score: 10
Sequence: 1 RM 2

Scoring table: BLOSUM62
Gapop 10.0 , Gapext 0.5

Searched: 572060 seqs, 82675679 residues

Total number of hits satisfying chosen parameters: 572060

Minimum DB seq length: 0
Maximum DB seq length: 2000000000

Post-processing: Minimum Match 0%
Maximum Match 100%
Listing first 45 summaries

Database : Issued_Patents_AA:*
1: /cgn2_6/ptodata/1/iaa/5_COMB.pep:*
2: /cgn2_6/ptodata/1/iaa/6_COMB.pep:*
3: /cgn2_6/ptodata/1/iaa/H_COMB.pep:*
4: /cgn2_6/ptodata/1/iaa/PCTUS_COMB.pep:*
5: /cgn2_6/ptodata/1/iaa/RE_COMB.pep:*
6: /cgn2_6/ptodata/1/iaa/backfiles1.pep:*

Pred. No. is the number of results predicted by chance to have a score greater than or equal to the score of the result being printed, and is derived by analysis of the total score distribution.

SUMMARIES

Result No.	Score	% Query Match	Length	DB	ID	Description
1	10	100.0	3	1	US-08-968-676-27	Sequence 27, Appl
2	10	100.0	3	2	US-09-261-894A-27	Sequence 27, Appl
3	10	100.0	4	1	US-08-331-383-4	Sequence 4, Appli
4	10	100.0	4	1	US-08-549-008-4	Sequence 4, Appli
5	10	100.0	4	1	US-08-968-676-26	Sequence 26, Appl
6	10	100.0	4	2	US-09-008-308-23	Sequence 23, Appl
7	10	100.0	4	2	US-09-191-906A-2	Sequence 2, Appli
8	10	100.0	4	2	US-09-101-649-12	Sequence 12, Appl
9	10	100.0	4	2	US-09-055-075C-16	Sequence 16, Appl
10	10	100.0	4	2	US-09-055-075C-18	Sequence 18, Appl
11	10	100.0	4	2	US-09-396-813-3	Sequence 3, Appli
12	10	100.0	4	2	US-09-919-124-16	Sequence 16, Appl
13	10	100.0	4	2	US-09-919-124-18	Sequence 18, Appl
14	10	100.0	4	2	US-09-680-201-2	Sequence 2, Appli
15	10	100.0	4	2	US-09-256-237-23	Sequence 23, Appl
16	10	100.0	4	2	US-09-261-894A-26	Sequence 26, Appl
17	10	100.0	4	2	US-09-763-293A-2	Sequence 2, Appli
18	10	100.0	4	2	US-09-763-293A-30	Sequence 30, Appl
19	10	100.0	4	2	US-09-763-293A-31	Sequence 31, Appl
20	10	100.0	4	2	US-09-763-293A-34	Sequence 34, Appl
21	10	100.0	4	4	PCT-US92-09070-2	Sequence 2, Appli
22	10	100.0	4	4	PCT-US92-09070-14	Sequence 14, Appl
23	10	100.0	4	4	PCT-US93-01669-9	Sequence 9, Appli
24	10	100.0	5	1	US-08-297-330-10	Sequence 10, Appl
25	10	100.0	5	1	US-08-504-265B-76	Sequence 76, Appl
26	10	100.0	5	1	US-08-894-922A-2	Sequence 2, Appli
27	10	100.0	5	1	US-08-968-676-25	Sequence 25, Appl
28	10	100.0	5	1	US-08-350-260A-413	Sequence 413, App
29	10	100.0	5	1	US-08-818-253-58	Sequence 58, Appl
30	10	100.0	5	2	US-08-894-173-46	Sequence 46, Appl
31	10	100.0	5	2	US-09-020-880-75	Sequence 75, Appl
32	10	100.0	5	2	US-09-020-880-77	Sequence 77, Appl
33	10	100.0	5	2	US-09-020-880-78	Sequence 78, Appl
34	10	100.0	5	2	US-08-974-549A-173	Sequence 173, App
35	10	100.0	5	2	US-09-398-193-46	Sequence 46, Appl
36	10	100.0	5	2	US-09-191-906A-9	Sequence 9, Appli
37	10	100.0	5	2	US-09-101-649-13	Sequence 13, Appl
38	10	100.0	5	2	US-08-930-921-4	Sequence 4, Appli
39	10	100.0	5	2	US-08-818-112-90	Sequence 90, Appl
40	10	100.0	5	2	US-08-818-111-91	Sequence 91, Appl
41	10	100.0	5	2	US-09-056-556-90	Sequence 90, Appl
42	10	100.0	5	2	US-09-101-544-75	Sequence 75, Appl
43	10	100.0	5	2	US-09-101-544-77	Sequence 77, Appl
44	10	100.0	5	2	US-09-101-544-78	Sequence 78, Appl
45	10	100.0	5	2	US-09-359-929-1	Sequence 1, Appli

ALIGNMENTS

```
RESULT 1
US-08-968-676-27
; Sequence 27, Application US/08968676
; Patent No. 5919639
;  GENERAL INFORMATION:
;    APPLICANT:  Humphreys, Robert E
;    APPLICANT:  Adams, Sharlene
;    APPLICANT:  Xu, Minzhen
;    TITLE OF INVENTION:  IMMUNOTHERAPY BY MODULATION OF ANTIGEN
;    NUMBER OF SEQUENCES:  165
;    CORRESPONDENCE ADDRESS:
;      ADDRESSEE:  Kevin M. Farrell, P.C.
;      STREET:  P.O. Box 999
;      CITY:  York Harbor
;      STATE:  ME
;      COUNTRY:  USA
;      ZIP:  03911
;    COMPUTER READABLE FORM:
;      MEDIUM TYPE:  Floppy disk
;      COMPUTER:  IBM PC compatible
;      OPERATING SYSTEM:  PC-DOS/MS-DOS
;      SOFTWARE:  PatentIn Release #1.0, Version #1.25
;    CURRENT APPLICATION DATA:
;      APPLICATION NUMBER:  US/08/968,676
;      FILING DATE:
;      CLASSIFICATION:  424
;    ATTORNEY/AGENT INFORMATION:
;      NAME:  Farrell, Kevin M
;      REGISTRATION NUMBER:  35,505
;      REFERENCE/DOCKET NUMBER:  REH-9601
;    TELECOMMUNICATION INFORMATION:
;      TELEPHONE:  (207) 363-0558
;      TELEFAX:  (207) 363-0528
;  INFORMATION FOR SEQ ID NO:  27:
;    SEQUENCE CHARACTERISTICS:
;      LENGTH:  3 amino acids
;      TYPE:  amino acid
;      STRANDEDNESS:  single
;      TOPOLOGY:  linear
;    MOLECULE TYPE:  peptide
US-08-968-676-27

  Query Match           100.0%;  Score 10;  DB 1;  Length 3;
  Best Local Similarity  100.0%;  Pred. No. 4.6e+05;
  Matches    2;  Conservative    0;  Mismatches    0;  Indels    0;  Gaps    0;

Qy          1 RM 2
               ||
Db          2 RM 3
```

```
RESULT 2
US-09-261-894A-27
; Sequence 27, Application US/09261894A
; Patent No. 6835382
; GENERAL INFORMATION:
;  APPLICANT: HUMPHREYS, ROBERT E.
;  APPLICANT:  ADAMS, SHARLENE
;  APPLICANT:  XU, MINZHEN
;  TITLE OF INVENTION: IMMUNOTHERAPY BY MODULATION OF ANTIGEN PRESENTATION
;  FILE REFERENCE: REH-2006
;  CURRENT APPLICATION NUMBER: US/09/261,894A
;  CURRENT FILING DATE:  1999-03-03
;  NUMBER OF SEQ ID NOS: 167
;  SOFTWARE: PatentIn Ver. 3.2
; SEQ ID NO 27
;   LENGTH: 3
;   TYPE: PRT
;   ORGANISM: Artificial Sequence
;   FEATURE:
;   OTHER INFORMATION: Description of Artificial Sequence: Synthetic
;   OTHER INFORMATION: peptide
US-09-261-894A-27

  Query Match             100.0%;  Score 10;  DB 2;  Length 3;
  Best Local Similarity   100.0%;  Pred. No. 4.6e+05;
  Matches    2;  Conservative    0;  Mismatches    0;  Indels    0;  Gaps    0;

Qy          1 RM 2
              ||
Db          2 RM 3


RESULT 3
US-08-331-383-4
; Sequence 4, Application US/08331383
; Patent No. 5605809
;  GENERAL INFORMATION:
;    APPLICANT:  Komoriya, Akira
;    APPLICANT:  Packard, Beverly S.
;    TITLE OF INVENTION:  Compositions for the Detection of
;    TITLE OF INVENTION:  Proteases in Biological Samples and Methods and Use
;    TITLE OF INVENTION:  Thereof
;    NUMBER OF SEQUENCES:  56
;    CORRESPONDENCE ADDRESS:
;      ADDRESSEE:  Townsend and Townsend Khourie and Crew
;      STREET:  One Market Plaza, Steuart Tower, Suite 2000
;      CITY:  San Francisco
;      STATE:  California
;      COUNTRY:  USA
;      ZIP:  94105
;    COMPUTER READABLE FORM:
;      MEDIUM TYPE:  Floppy disk
;      COMPUTER:  IBM PC compatible
;      OPERATING SYSTEM:  PC-DOS/MS-DOS
;      SOFTWARE:  PatentIn Release #1.0, Version #1.25
;    CURRENT APPLICATION DATA:
;      APPLICATION NUMBER:  US/08/331,383
;      FILING DATE:  28-OCT-1994
;      CLASSIFICATION:  424
;    ATTORNEY/AGENT INFORMATION:
;      NAME:  Dow, Karen B.
;      REGISTRATION NUMBER:  29,684
;      REFERENCE/DOCKET NUMBER:  16865-1
;    TELECOMMUNICATION INFORMATION:
;      TELEPHONE:  415-326-2400
;      TELEFAX:  415-326-2422
;  INFORMATION FOR SEQ ID NO:  4:
;    SEQUENCE CHARACTERISTICS:
;      LENGTH:  4 amino acids
;      TYPE:  amino acid
;      STRANDEDNESS:  single
;      TOPOLOGY:  linear
;    MOLECULE TYPE:  peptide
US-08-331-383-4

  Query Match             100.0%;  Score 10;  DB 1;  Length 4;
  Best Local Similarity   100.0%;  Pred. No. 4.6e+05;
  Matches    2;  Conservative    0;  Mismatches    0;  Indels    0;  Gaps    0;

Qy          1 RM 2
              ||
Db          1 RM 2


RESULT 4
US-08-549-008-4
; Sequence 4, Application US/08549008
; Patent No. 5714342
;  GENERAL INFORMATION:
;    APPLICANT:  Komoriya, Akira
;    APPLICANT:  Packard, Beverly S.
;    TITLE OF INVENTION:  Compositions for the Detection of
;    TITLE OF INVENTION:  Protease in Biological Samples and Methods of Use Thereof
;    NUMBER OF SEQUENCES:  56
;    CORRESPONDENCE ADDRESS:
;      ADDRESSEE:  Townsend and Townsend and Crew LLP
;      STREET:  Two Embarcadero Center, Eighth Floor
;      CITY:  San Francisco
;      STATE:  California
;      COUNTRY:  USA
;      ZIP:  94111-3834
;    COMPUTER READABLE FORM:
;      MEDIUM TYPE:  Floppy disk
;      COMPUTER:  IBM PC compatible
;      OPERATING SYSTEM:  PC-DOS/MS-DOS
;      SOFTWARE:  PatentIn Release #1.0, Version #1.30
;    CURRENT APPLICATION DATA:
;      APPLICATION NUMBER:  US/08/549,008
;      FILING DATE:  27-OCT-1995
;      CLASSIFICATION:  435
;    PRIOR APPLICATION DATA:
;      APPLICATION NUMBER:  US 08/331,383
;      FILING DATE:  28-OCT-1994
;    ATTORNEY/AGENT INFORMATION:
;      NAME:  Weber, Ellen Lauver
;      REGISTRATION NUMBER:  32,762
;      REFERENCE/DOCKET NUMBER:  016865-000110US
;    TELECOMMUNICATION INFORMATION:
;      TELEPHONE:  (415) 576-0200
;      TELEFAX:  (415) 576-0300
;  INFORMATION FOR SEQ ID NO:  4:
;    SEQUENCE CHARACTERISTICS:
;      LENGTH:  4 amino acids
;      TYPE:  amino acid
;      STRANDEDNESS:
;      TOPOLOGY:  linear
;    MOLECULE TYPE:  peptide
US-08-549-008-4

  Query Match             100.0%;  Score 10;  DB 1;  Length 4;
  Best Local Similarity   100.0%;  Pred. No. 4.6e+05;
  Matches    2;  Conservative    0;  Mismatches    0;  Indels    0;  Gaps    0;

Qy          1 RM 2
              ||
Db          1 RM 2


RESULT 5
US-08-968-676-26
; Sequence 26, Application US/08968676
; Patent No. 5919639
;  GENERAL INFORMATION:
;    APPLICANT:  Humphreys, Robert E
;    APPLICANT:  Adams, Sharlene
```

```
;    APPLICANT:  Xu, Minzhen
;    TITLE OF INVENTION:  IMMUNOTHERAPY BY MODULATION OF ANTIGEN
;    NUMBER OF SEQUENCES:  165
;    CORRESPONDENCE ADDRESS:
;      ADDRESSEE:  Kevin M. Farrell, P.C.
;      STREET:  P.O. Box 999
;      CITY:  York Harbor
;      STATE:  ME
;      COUNTRY:  USA
;      ZIP:  03911
;    COMPUTER READABLE FORM:
;      MEDIUM TYPE:  Floppy disk
;      COMPUTER:  IBM PC compatible
;      OPERATING SYSTEM:  PC-DOS/MS-DOS
;      SOFTWARE:  PatentIn Release #1.0, Version #1.25
;    CURRENT APPLICATION DATA:
;      APPLICATION NUMBER:  US/08/968,676
;      FILING DATE:
;      CLASSIFICATION:  424
;    ATTORNEY/AGENT INFORMATION:
;      NAME:  Farrell, Kevin M
;      REGISTRATION NUMBER:  35,505
;      REFERENCE/DOCKET NUMBER:  REH-9601
;    TELECOMMUNICATION INFORMATION:
;      TELEPHONE:  (207) 363-0558
;      TELEFAX:  (207) 363-0528
;  INFORMATION FOR SEQ ID NO:  26:
;    SEQUENCE CHARACTERISTICS:
;      LENGTH:  4 amino acids
;      TYPE:  amino acid
;      STRANDEDNESS:  single
;      TOPOLOGY:  linear
;    MOLECULE TYPE:  peptide
US-08-968-676-26

  Query Match             100.0%;  Score 10;  DB 1;  Length 4;
  Best Local Similarity   100.0%;  Pred. No. 4.6e+05;
  Matches    2;  Conservative    0;  Mismatches    0;  Indels    0;  Gaps    0;

Qy          1 RM 2
              ||
Db          2 RM 3


RESULT 6
US-09-008-308-23
; Sequence 23, Application US/09008308
; Patent No. 6080575
;  GENERAL INFORMATION:
;    APPLICANT:  Heidtmann, Hans H.
;    APPLICANT:  Mueller, Rolf
;    APPLICANT:  Sedlacek, Hans-Harald
;    TITLE OF INVENTION:  NUCLEIC ACID CONSTRUCT FOR EXPRESSING
;    TITLE OF INVENTION:  ACTIVE SUBSTANCES WHICH CAN BE ACTIVATED BY PROTEASES, AND
;    TITLE OF INVENTION:  PREPARATION AND USE
;    NUMBER OF SEQUENCES:  62
;    CORRESPONDENCE ADDRESS:
;      ADDRESSEE:  Foley & Lardner
;      STREET:  3000 K Street, N.W., Suite 500
;      CITY:  Washington
;      STATE:  D.C.
;      COUNTRY:  USA
;      ZIP:  20007-5109
;    COMPUTER READABLE FORM:
;      MEDIUM TYPE:  Floppy disk
;      COMPUTER:  IBM PC compatible
;      OPERATING SYSTEM:  PC-DOS/MS-DOS
;      SOFTWARE:  PatentIn Release #1.0, Version #1.30
;    CURRENT APPLICATION DATA:
;      APPLICATION NUMBER:  US/09/008,308
;      FILING DATE:  16-JAN-1998
;      CLASSIFICATION:  514
;    PRIOR APPLICATION DATA:
;      APPLICATION NUMBER:  DE 197 01 141.1
;      FILING DATE:  16-JAN-1997
;    ATTORNEY/AGENT INFORMATION:
;      NAME:  Sandercock, Colin G.
;      REGISTRATION NUMBER:  31,298
;      REFERENCE/DOCKET NUMBER:  026083/0189
;    TELECOMMUNICATION INFORMATION:
;      TELEPHONE:  (202)672-5300
;      TELEFAX:  (202)672-5399
;      TELEX:  904136
;  INFORMATION FOR SEQ ID NO:  23:
;    SEQUENCE CHARACTERISTICS:
;      LENGTH:  4 amino acids
;      TYPE:  amino acid
;      STRANDEDNESS:
;      TOPOLOGY:  linear
;    MOLECULE TYPE:  peptide
;    FRAGMENT TYPE:  internal
US-09-008-308-23

  Query Match             100.0%;  Score 10;  DB 2;  Length 4;
  Best Local Similarity   100.0%;  Pred. No. 4.6e+05;
  Matches    2;  Conservative    0;  Mismatches    0;  Indels    0;  Gaps    0;

Qy          1 RM 2
              ||
Db          3 RM 4


RESULT 7
US-09-191-906A-2
; Sequence 2, Application US/09191906A
; Patent No. 6207643
; GENERAL INFORMATION:
;  APPLICANT:  Nachman, Ronald J
;  APPLICANT:  Garside, Christopher S
;  APPLICANT:  Tobe, Stephen S
;  TITLE OF INVENTION:  Mimetic Insect Allatostatin Analogs for Insect Control
;  FILE REFERENCE:  P.C. 0047.99-Ronald J. Nachman et al.
;  CURRENT APPLICATION NUMBER: US/09/191,906A
;  CURRENT FILING DATE:  1998-11-13
;  NUMBER OF SEQ ID NOS:  14
;  SOFTWARE:  PatentIn Ver. 2.1
; SEQ ID NO 2
;   LENGTH: 4
;   TYPE: PRT
;   ORGANISM: Diploptera punctata
US-09-191-906A-2

  Query Match             100.0%;  Score 10;  DB 2;  Length 4;
  Best Local Similarity   100.0%;  Pred. No. 4.6e+05;
  Matches    2;  Conservative    0;  Mismatches    0;  Indels    0;  Gaps    0;

Qy          1 RM 2
              ||
Db          3 RM 4


RESULT 8
US-09-101-649-12
; Sequence 12, Application US/09101649
; Patent No. 6262017
;   GENERAL INFORMATION:
;        APPLICANT: Dee, Kay C
;                   Andersen, Thomas T.
;                   Bizios, Rena
;        TITLE OF INVENTION: Peptides for
;                            Altering Osteoblast Adhesion
;        NUMBER OF SEQUENCES: 53
;        CORRESPONDENCE ADDRESS:
;             ADDRESSEE: Heslin & Rothenberg, P.C.
;             STREET: 5 Columbia Circle
;             CITY: Albany
```

; STATE: New York
; COUNTRY: USA
; ZIP: 12203
; COMPUTER READABLE FORM:
; MEDIUM TYPE: Diskette, 3.5 inch, 1.44 MB storage
; COMPUTER: DTK Grafika 486
; OPERATING SYSTEM: MS-DOS 5.0
; SOFTWARE: Word Perfect 5.1
; CURRENT APPLICATION DATA:
; APPLICATION NUMBER: US/09/101,649
; FILING DATE: 14-Jul-1998
; PRIOR APPLICATION DATA:
; APPLICATION NUMBER: 60/010,026
; FILING DATE: 16-JAN-1996
; APPLICATION NUMBER: 60/029,189
; FILING DATE: 31-OCT-1996
; ATTORNEY/AGENT INFORMATION:
; NAME: Hansen, Philip E.
; REGISTRATION NUMBER: 32700
; REFERENCE/DOCKET NUMBER: 0094.026BWO
; TELECOMMUNICATION INFORMATION:
; TELEPHONE: (518) 452-5600
; TELEFAX: (518) 452-5579
; INFORMATION FOR SEQ ID NO: 12
; SEQUENCE CHARACTERISTICS:
; LENGTH: 4 amino acids
; TYPE: amino acid
; TOPOLOGY: linear
; MOLECULE TYPE: peptide
; HYPOTHETICAL: yes
; SEQUENCE DESCRIPTION: SEQ ID NO: 12
US-09-101-649-12

Query Match 100.0%; Score 10; DB 2; Length 4;
Best Local Similarity 100.0%; Pred. No. 4.6e+05;
Matches 2; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

Qy 1 RM 2
||
Db 2 RM 3

RESULT 9
US-09-055-075C-16
; Sequence 16, Application US/09055075C
; Patent No. 6328976
; GENERAL INFORMATION:
; APPLICANT: Luftig, Ronald B.
; TITLE OF INVENTION: No. 6328976-Infectious, Protease Defective HIV Particles and
; TITLE OF INVENTION: Nucleic Acid Molecules Encoding Therefor
; FILE REFERENCE: P-IM 3074
; CURRENT APPLICATION NUMBER: US/09/055,075C
; CURRENT FILING DATE: 1998-04-03
; PRIOR APPLICATION NUMBER: 60/043,047
; PRIOR FILING DATE: 1997-04-04
; NUMBER OF SEQ ID NOS: 48
; SOFTWARE: PatentIn Ver. 2.0
; SEQ ID NO 16
; LENGTH: 4
; TYPE: PRT
; ORGANISM: Human immunodeficiency virus type 1
US-09-055-075C-16

Query Match 100.0%; Score 10; DB 2; Length 4;
Best Local Similarity 100.0%; Pred. No. 4.6e+05;
Matches 2; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

Qy 1 RM 2
||
Db 2 RM 3

;
RESULT 10
US-09-055-075C-18
; Sequence 18, Application US/09055075C
; Patent No. 6328976
; GENERAL INFORMATION:
; APPLICANT: Luftig, Ronald B.
; TITLE OF INVENTION: No. 6328976-Infectious, Protease Defective HIV Particles and
; TITLE OF INVENTION: Nucleic Acid Molecules Encoding Therefor
; FILE REFERENCE: P-IM 3074
; CURRENT APPLICATION NUMBER: US/09/055,075C
; CURRENT FILING DATE: 1998-04-03
; PRIOR APPLICATION NUMBER: 60/043,047
; PRIOR FILING DATE: 1997-04-04
; NUMBER OF SEQ ID NOS: 48
; SOFTWARE: PatentIn Ver. 2.0
; SEQ ID NO 18
; LENGTH: 4
; TYPE: PRT
; ORGANISM: Human immunodeficiency virus type 1
US-09-055-075C-18

Query Match 100.0%; Score 10; DB 2; Length 4;
Best Local Similarity 100.0%; Pred. No. 4.6e+05;
Matches 2; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

Qy 1 RM 2
||
Db 2 RM 3

RESULT 11
US-09-396-813-3
; Sequence 3, Application US/09396813
; Patent No. 6432409
; GENERAL INFORMATION:
; APPLICANT: Humphreys, Robert E.
; APPLICANT: Adams, Sharlene
; APPLICANT: Xu, Minzhen
; TITLE OF INVENTION: HYBRID PEPTIDES MODULATE THE IMMUNE RESPONSE
; FILE REFERENCE: REH2007
; CURRENT APPLICATION NUMBER: US/09/396,813
; CURRENT FILING DATE: 1999-09-14
; NUMBER OF SEQ ID NOS: 12
; SOFTWARE: PatentIn Ver. 2.0
; SEQ ID NO 3
; LENGTH: 4
; TYPE: PRT
; ORGANISM: Artificial Sequence
; FEATURE:
; OTHER INFORMATION: Description of Artificial Sequence: Modified mouse
; OTHER INFORMATION: Ii key peptide
US-09-396-813-3

Query Match 100.0%; Score 10; DB 2; Length 4;
Best Local Similarity 100.0%; Pred. No. 4.6e+05;
Matches 2; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

Qy 1 RM 2
||
Db 2 RM 3

RESULT 12
US-09-919-124-16
; Sequence 16, Application US/09919124
; Patent No. 6557296
; GENERAL INFORMATION:
; APPLICANT: Luftig, Ronald B.
; TITLE OF INVENTION: No. 6557296-Infectious, Protease Defective HIV Particles and
; TITLE OF INVENTION: Nucleic Acid Molecules Encoding Therefor
; FILE REFERENCE: P-IM 4867
; CURRENT APPLICATION NUMBER: US/09/919,124
; CURRENT FILING DATE: 2001-07-30
; PRIOR APPLICATION NUMBER: 60/043,047

```
;  PRIOR FILING DATE: 1997-04-04
;  PRIOR APPLICATION NUMBER: US 09/055,075
;  PRIOR FILING DATE: 1998-04-03
;  NUMBER OF SEQ ID NOS: 48
;  SOFTWARE: PatentIn Ver. 2.0
; SEQ ID NO 16
;   LENGTH: 4
;   TYPE: PRT
;   ORGANISM: Human immunodeficiency virus type 1
US-09-919-124-16

  Query Match           100.0%;  Score 10;  DB 2;  Length 4;
  Best Local Similarity  100.0%;  Pred. No. 4.6e+05;
  Matches    2;  Conservative    0;  Mismatches    0;  Indels    0;  Gaps    0;

Qy          1 RM 2
              ||
Db          2 RM 3


RESULT 13
US-09-919-124-18
; Sequence 18, Application US/09919124
; Patent No. 6557296
; GENERAL INFORMATION:
;  APPLICANT: Luftig, Ronald B.
;  TITLE OF INVENTION: No. 6557296-Infectious, Protease Defective HIV Particles and
;  TITLE OF INVENTION:  Nucleic Acid Molecules Encoding Therefor
;  FILE REFERENCE: P-IM 4867
;  CURRENT APPLICATION NUMBER: US/09/919,124
;  CURRENT FILING DATE:  2001-07-30
;  PRIOR APPLICATION NUMBER: 60/043,047
;  PRIOR FILING DATE: 1997-04-04
;  PRIOR APPLICATION NUMBER: US 09/055,075
;  PRIOR FILING DATE: 1998-04-03
;  NUMBER OF SEQ ID NOS: 48
;  SOFTWARE: PatentIn Ver. 2.0
; SEQ ID NO 18
;   LENGTH: 4
;   TYPE: PRT
;   ORGANISM: Human immunodeficiency virus type 1
US-09-919-124-18

  Query Match           100.0%;  Score 10;  DB 2;  Length 4;
  Best Local Similarity  100.0%;  Pred. No. 4.6e+05;
  Matches    2;  Conservative    0;  Mismatches    0;  Indels    0;  Gaps    0;

Qy          1 RM 2
              ||
Db          2 RM 3


RESULT 14
US-09-680-201-2
; Sequence 2, Application US/09680201
; Patent No. 6664371
; GENERAL INFORMATION:
;  APPLICANT: Nachman, Ronald J
;  APPLICANT:  Garside, Christopher S
;  APPLICANT:  Tobe, Stephen S
;  TITLE OF INVENTION: Mimetic Insect Allatostatin Analogs for Insect Control
;  FILE REFERENCE: P.C. 0047.99-Ronald J. Nachman et al.
;  CURRENT APPLICATION NUMBER: US/09/680,201
;  CURRENT FILING DATE:  2001-01-19
;  NUMBER OF SEQ ID NOS: 14
;  SOFTWARE: PatentIn Ver. 2.1
; SEQ ID NO 2
;   LENGTH: 4
;   TYPE: PRT
;   ORGANISM: Diploptera punctata
US-09-680-201-2

  Query Match           100.0%;  Score 10;  DB 2;  Length 4;
  Best Local Similarity  100.0%;  Pred. No. 4.6e+05;
  Matches    2;  Conservative    0;  Mismatches    0;  Indels    0;  Gaps    0;

Qy          1 RM 2
              ||
Db          3 RM 4


RESULT 15
US-09-256-237-23
; Sequence 23, Application US/09256237
; Patent No. 6670147
;   GENERAL INFORMATION:
;        APPLICANT: Heidtmann, Hans H.
;                   Mueller, Rolf
;                   Sedlacek, Hans-Harald
;        TITLE OF INVENTION: NUCLEIC ACID CONSTRUCT FOR EXPRESSING
;                            ACTIVE SUBSTANCES WHICH CAN BE ACTIVATED BY PROTEASES, AND
;                            PREPARATION AND USE
;        NUMBER OF SEQUENCES: 62
;        CORRESPONDENCE ADDRESS:
;             ADDRESSEE: Foley & Lardner
;             STREET: 3000 K Street, N.W., Suite 500
;             CITY: Washington
;             STATE: D.C.
;             COUNTRY: USA
;             ZIP: 20007-5109
;        COMPUTER READABLE FORM:
;             MEDIUM TYPE: Floppy disk
;             COMPUTER: IBM PC compatible
;             OPERATING SYSTEM: PC-DOS/MS-DOS
;             SOFTWARE: PatentIn Release #1.0, Version #1.30
;        CURRENT APPLICATION DATA:
;             APPLICATION NUMBER: US/09/256,237
;             FILING DATE: 24-Feb-1999
;             CLASSIFICATION: <Unknown>
;        PRIOR APPLICATION DATA:
;             APPLICATION NUMBER: US/09/008,308
;             FILING DATE: 16-JAN-1998
;             APPLICATION NUMBER: DE 197 01 141.1
;             FILING DATE: 16-JAN-1997
;        ATTORNEY/AGENT INFORMATION:
;             NAME: Sandercock, Colin G.
;             REGISTRATION NUMBER: 31,298
;             REFERENCE/DOCKET NUMBER: 026083/0189
;        TELECOMMUNICATION INFORMATION:
;             TELEPHONE: (202)672-5300
;             TELEFAX: (202)672-5399
;             TELEX: 904136
;   INFORMATION FOR SEQ ID NO: 23:
;        SEQUENCE CHARACTERISTICS:
;             LENGTH: 4 amino acids
;             TYPE: amino acid
;             STRANDEDNESS: <Unknown>
;             TOPOLOGY: linear
;        MOLECULE TYPE: peptide
;        FRAGMENT TYPE: internal
;        SEQUENCE DESCRIPTION: SEQ ID NO: 23:
US-09-256-237-23

  Query Match           100.0%;  Score 10;  DB 2;  Length 4;
  Best Local Similarity  100.0%;  Pred. No. 4.6e+05;
  Matches    2;  Conservative    0;  Mismatches    0;  Indels    0;  Gaps    0;

Qy          1 RM 2
              ||
Db          3 RM 4


Search completed: February 28, 2006, 08:56:08
Job time : 1.4221 secs
```

This Page Blank (uspto)

GenCore version 5.1.7

OM protein - protein search, using sw model

Run on: February 28, 2006, 08:54:36 ; Search time 0.583124 Seconds
(without alignments)
1433.071 Million cell updates/sec

Title: US-10-717-665A-44_COPY_448_449
Perfect score: 10
Sequence: 1 RM 2

Scoring table: BLOSUM62
Gapop 10.0 , Gapext 0.5

Searched: 1867569 seqs, 417829326 residues

Total number of hits satisfying chosen parameters: 1867569

Minimum DB seq length: 0
Maximum DB seq length: 2000000000

Post-processing: Minimum Match 0%
Maximum Match 100%
Listing first 45 summaries

Database : Published_Applications_AA_Main:*
1: /cgn2_6/ptodata/1/pubpaa/US07_PUBCOMB.pep:*
2: /cgn2_6/ptodata/1/pubpaa/US08_PUBCOMB.pep:*
3: /cgn2_6/ptodata/1/pubpaa/US09_PUBCOMB.pep:*
4: /cgn2_6/ptodata/1/pubpaa/US10A_PUBCOMB.pep:*
5: /cgn2_6/ptodata/1/pubpaa/US10B_PUBCOMB.pep:*
6: /cgn2_6/ptodata/1/pubpaa/US11_PUBCOMB.pep:*

Pred. No. is the number of results predicted by chance to have a score greater than or equal to the score of the result being printed, and is derived by analysis of the total score distribution.

SUMMARIES

Result No.	Score	% Query Match	Length	DB	ID	Description
1	10	100.0	3	3	US-09-261-894-27	Sequence 27, Appl
2	10	100.0	4	2	US-08-484-409-31	Sequence 31, Appl
3	10	100.0	4	3	US-09-919-124-16	Sequence 16, Appl
4	10	100.0	4	3	US-09-919-124-18	Sequence 18, Appl
5	10	100.0	4	3	US-09-998-491-10	Sequence 10, Appl
6	10	100.0	4	3	US-09-261-894-26	Sequence 26, Appl
7	10	100.0	4	4	US-10-032-717-48	Sequence 48, Appl
8	10	100.0	4	4	US-10-197-000-3	Sequence 3, Appli
9	10	100.0	4	4	US-10-163-198-86	Sequence 86, Appl
10	10	100.0	4	4	US-10-385-317-2	Sequence 2, Appli
11	10	100.0	4	4	US-10-414-637-48	Sequence 48, Appl
12	10	100.0	4	4	US-10-245-871-3	Sequence 3, Appli
13	10	100.0	4	4	US-10-245-871-8	Sequence 8, Appli
14	10	100.0	4	4	US-10-245-871-801	Sequence 801, App
15	10	100.0	4	4	US-10-659-509-2	Sequence 2, Appli
16	10	100.0	4	4	US-10-253-286-3	Sequence 3, Appli
17	10	100.0	4	4	US-10-253-286-8	Sequence 8, Appli
18	10	100.0	4	4	US-10-253-286-801	Sequence 801, App
19	10	100.0	4	4	US-10-475-281-10	Sequence 10, Appl
20	10	100.0	4	5	US-10-659-233-2	Sequence 2, Appli
21	10	100.0	4	5	US-10-994-727-10	Sequence 10, Appl
22	10	100.0	5	3	US-09-096-749A-53	Sequence 53, Appl
23	10	100.0	5	3	US-09-873-459A-30	Sequence 30, Appl
24	10	100.0	5	3	US-09-873-459A-31	Sequence 31, Appl
25	10	100.0	5	3	US-09-192-854-45	Sequence 45, Appl
26	10	100.0	5	3	US-09-192-854-49	Sequence 49, Appl
27	10	100.0	5	3	US-09-192-854-52	Sequence 52, Appl
28	10	100.0	5	3	US-09-192-854-77	Sequence 77, Appl
29	10	100.0	5	3	US-09-192-854-108	Sequence 108, App
30	10	100.0	5	3	US-09-192-854-120	Sequence 120, App
31	10	100.0	5	3	US-09-751-100B-46	Sequence 46, Appl
32	10	100.0	5	3	US-09-968-561A-67	Sequence 67, Appl
33	10	100.0	5	3	US-09-968-561A-73	Sequence 73, Appl
34	10	100.0	5	3	US-09-968-561A-79	Sequence 79, Appl
35	10	100.0	5	3	US-09-968-561A-85	Sequence 85, Appl
36	10	100.0	5	3	US-09-968-561A-91	Sequence 91, Appl
37	10	100.0	5	3	US-09-968-561A-127	Sequence 127, App
38	10	100.0	5	3	US-09-968-561A-139	Sequence 139, App
39	10	100.0	5	3	US-09-968-561A-145	Sequence 145, App
40	10	100.0	5	3	US-09-968-561A-187	Sequence 187, App
41	10	100.0	5	3	US-09-968-561A-211	Sequence 211, App
42	10	100.0	5	3	US-09-968-561A-217	Sequence 217, App
43	10	100.0	5	3	US-09-903-412-53	Sequence 53, Appl
44	10	100.0	5	3	US-09-781-796B-4	Sequence 4, Appli
45	10	100.0	5	3	US-09-968-744A-67	Sequence 67, Appl

ALIGNMENTS

```
RESULT 1
US-09-261-894-27
; Sequence 27, Application US/09261894
; Publication No. US20030207324A1
;  GENERAL INFORMATION:
;    APPLICANT:  Humphreys, Robert E
;    APPLICANT:  Adams, Sharlene
;    APPLICANT:  Xu, Minzhen
;    TITLE OF INVENTION:  IMMUNOTHERAPY BY MODULATION OF ANTIGEN
;    NUMBER OF SEQUENCES:  165
;    CORRESPONDENCE ADDRESS:
;      ADDRESSEE:  Kevin M. Farrell, P.C.
;      STREET:  P.O. Box 999
;      CITY:  York Harbor
;      STATE:  ME
;      COUNTRY:  USA
;      ZIP:  03911
;    COMPUTER READABLE FORM:
;      MEDIUM TYPE:  Floppy disk
;      COMPUTER:  IBM PC compatible
;      OPERATING SYSTEM:  PC-DOS/MS-DOS
;      SOFTWARE:  PatentIn Release #1.0, Version #1.25
;    CURRENT APPLICATION DATA:
;      APPLICATION NUMBER:  US/09/261,894
;      FILING DATE:  March 3, 1999
;      CLASSIFICATION:
;    ATTORNEY/AGENT INFORMATION:
;      NAME:  Farrell, Kevin M
;      REGISTRATION NUMBER:  35,505
;      REFERENCE/DOCKET NUMBER:  REH-9601
;    TELECOMMUNICATION INFORMATION:
;      TELEPHONE:  (207) 363-0558
;      TELEFAX:  (207) 363-0528
;  INFORMATION FOR SEQ ID NO:  27:
;    SEQUENCE CHARACTERISTICS:
;      LENGTH:  3 amino acids
;      TYPE:  amino acid
;      STRANDEDNESS:  single
;      TOPOLOGY:  linear
;    MOLECULE TYPE:  peptide
US-09-261-894-27

  Query Match           100.0%;  Score 10;  DB 3;  Length 3;
  Best Local Similarity  100.0%;  Pred. No. 1.7e+06;
  Matches    2;  Conservative    0;  Mismatches    0;  Indels    0;  Gaps    0;

Qy          1 RM 2
               ||
Db          2 RM 3
```

```
RESULT 2
US-08-484-409-31
; Sequence 31, Application US/08484409
; Publication No. US20020076412A1
;  GENERAL INFORMATION:
;    APPLICANT:  Steinman, Lawrence
;    APPLICANT:  Zamvil, Scott
;    TITLE OF INVENTION:  METHODS FOR MODULATING THE IMMUNE SYSTEM
;    NUMBER OF SEQUENCES:  52
;    CORRESPONDENCE ADDRESS:
;      ADDRESSEE:  SEED and BERRY LLP
;      STREET:  6300 Columbia Center, 701 Fifth Avenue
;      CITY:  Seattle
;      STATE:  Washington
;      COUNTRY:  USA
;      ZIP:  98104-7092
;    COMPUTER READABLE FORM:
;      MEDIUM TYPE:  Floppy disk
;      COMPUTER:  IBM PC compatible
;      OPERATING SYSTEM:  PC-DOS/MS-DOS
;      SOFTWARE:  PatentIn Release #1.0, Version #1.30
;    CURRENT APPLICATION DATA:
;      APPLICATION NUMBER:  US/08/484,409
;      FILING DATE:  07-JUN-1995
;      CLASSIFICATION:  424
;    ATTORNEY/AGENT INFORMATION:
;      NAME:  Maki, David J.
;      REGISTRATION NUMBER:  31,392
;      REFERENCE/DOCKET NUMBER:  690068.409C1
;    TELECOMMUNICATION INFORMATION:
;      TELEPHONE:  (206) 622-4900
;      TELEFAX:  (206) 682-6031
;  INFORMATION FOR SEQ ID NO:  31:
;    SEQUENCE CHARACTERISTICS:
;      LENGTH:  4 amino acids
;      TYPE:  amino acid
;      STRANDEDNESS:
;      TOPOLOGY:  linear
US-08-484-409-31

  Query Match             100.0%;  Score 10;  DB 2;  Length 4;
  Best Local Similarity   100.0%;  Pred. No. 1.7e+06;
  Matches    2;  Conservative    0;  Mismatches    0;  Indels    0;  Gaps    0;

Qy          1 RM 2
              ||
Db          1 RM 2


RESULT 3
US-09-919-124-16
; Sequence 16, Application US/09919124
; Patent No. US20020061313A1
; GENERAL INFORMATION:
;  APPLICANT: Luftig, Ronald B.
;  TITLE OF INVENTION: No. US20020061313A1-Infectious, Protease Defective HIV Particles
;  TITLE OF INVENTION:  Nucleic Acid Molecules Encoding Therefor
;  FILE REFERENCE: P-IM 4867
;  CURRENT APPLICATION NUMBER: US/09/919,124
;  CURRENT FILING DATE:  2001-07-30
;  PRIOR APPLICATION NUMBER: 60/043,047
;  PRIOR FILING DATE: 1997-04-04
;  PRIOR APPLICATION NUMBER: US 09/055,075
;  PRIOR FILING DATE: 1998-04-03
;  NUMBER OF SEQ ID NOS: 48
;  SOFTWARE: PatentIn Ver. 2.0
; SEQ ID NO 16
;   LENGTH: 4
;   TYPE: PRT
;   ORGANISM: Human immunodeficiency virus type 1
US-09-919-124-16

  Query Match             100.0%;  Score 10;  DB 3;  Length 4;
  Best Local Similarity   100.0%;  Pred. No. 1.7e+06;
  Matches    2;  Conservative    0;  Mismatches    0;  Indels    0;  Gaps    0;

Qy          1 RM 2
              ||
Db          2 RM 3


RESULT 4
US-09-919-124-18
; Sequence 18, Application US/09919124
; Patent No. US20020061313A1
; GENERAL INFORMATION:
;  APPLICANT: Luftig, Ronald B.
;  TITLE OF INVENTION: No. US20020061313A1-Infectious, Protease Defective HIV Particles a
;  TITLE OF INVENTION:  Nucleic Acid Molecules Encoding Therefor
;  FILE REFERENCE: P-IM 4867
;  CURRENT APPLICATION NUMBER: US/09/919,124
;  CURRENT FILING DATE:  2001-07-30
;  PRIOR APPLICATION NUMBER: 60/043,047
;  PRIOR FILING DATE: 1997-04-04
;  PRIOR APPLICATION NUMBER: US 09/055,075
;  PRIOR FILING DATE: 1998-04-03
;  NUMBER OF SEQ ID NOS: 48
;  SOFTWARE: PatentIn Ver. 2.0
; SEQ ID NO 18
;   LENGTH: 4
;   TYPE: PRT
;   ORGANISM: Human immunodeficiency virus type 1
US-09-919-124-18

  Query Match             100.0%;  Score 10;  DB 3;  Length 4;
  Best Local Similarity   100.0%;  Pred. No. 1.7e+06;
  Matches    2;  Conservative    0;  Mismatches    0;  Indels    0;  Gaps    0;

Qy          1 RM 2
              ||
Db          2 RM 3


RESULT 5
US-09-998-491-10
; Sequence 10, Application US/09998491
; Publication No. US20030166529A1
; GENERAL INFORMATION:
;  APPLICANT: Mileusnic, Radmilla
;  APPLICANT:  Rose, Stephen Peter Russell
;  TITLE OF INVENTION: Polypeptides and their Uses
;  FILE REFERENCE: 3578-120
;  CURRENT APPLICATION NUMBER: US/09/998,491
;  CURRENT FILING DATE:  2001-11-30
;  PRIOR APPLICATION NUMBER: GB 0109558.7
;  PRIOR FILING DATE: 2001-04-18
;  PRIOR APPLICATION NUMBER: GB 0120084
;  PRIOR FILING DATE: 2001-08-07
;  NUMBER OF SEQ ID NOS: 11
;  SOFTWARE: FastSEQ for Windows Version 4.0
; SEQ ID NO 10
;   LENGTH: 4
;   TYPE: PRT
;   ORGANISM: Artificial Sequence
;   FEATURE:
;   OTHER INFORMATION: 4-mer polypeptide
US-09-998-491-10

  Query Match             100.0%;  Score 10;  DB 3;  Length 4;
  Best Local Similarity   100.0%;  Pred. No. 1.7e+06;
  Matches    2;  Conservative    0;  Mismatches    0;  Indels    0;  Gaps    0;

Qy          1 RM 2
              ||
Db          3 RM 4
```

```
RESULT 6
US-09-261-894-26
; Sequence 26, Application US/09261894
; Publication No. US20030207324A1
;  GENERAL INFORMATION:
;     APPLICANT:  Humphreys, Robert E
;     APPLICANT:  Adams, Sharlene
;     APPLICANT:  Xu, Minzhen
;     TITLE OF INVENTION:  IMMUNOTHERAPY BY MODULATION OF ANTIGEN
;     NUMBER OF SEQUENCES:  165
;     CORRESPONDENCE ADDRESS:
;       ADDRESSEE:  Kevin M. Farrell, P.C.
;       STREET:  P.O. Box 999
;       CITY:  York Harbor
;       STATE:  ME
;       COUNTRY:  USA
;       ZIP:  03911
;     COMPUTER READABLE FORM:
;       MEDIUM TYPE:  Floppy disk
;       COMPUTER:  IBM PC compatible
;       OPERATING SYSTEM:  PC-DOS/MS-DOS
;       SOFTWARE:  PatentIn Release #1.0, Version #1.25
;     CURRENT APPLICATION DATA:
;       APPLICATION NUMBER:  US/09/261,894
;       FILING DATE:  March 3, 1999
;       CLASSIFICATION:
;     ATTORNEY/AGENT INFORMATION:
;       NAME:  Farrell, Kevin M
;       REGISTRATION NUMBER:  35,505
;       REFERENCE/DOCKET NUMBER:  REH-9601
;     TELECOMMUNICATION INFORMATION:
;       TELEPHONE:  (207) 363-0558
;       TELEFAX:  (207) 363-0528
;   INFORMATION FOR SEQ ID NO:  26:
;     SEQUENCE CHARACTERISTICS:
;       LENGTH:  4 amino acids
;       TYPE:  amino acid
;       STRANDEDNESS:  single
;       TOPOLOGY:  linear
;     MOLECULE TYPE:  peptide
US-09-261-894-26

  Query Match             100.0%;  Score 10;  DB 3;  Length 4;
  Best Local Similarity   100.0%;  Pred. No. 1.7e+06;
  Matches    2;  Conservative    0;  Mismatches    0;  Indels    0;  Gaps    0;

Qy          1 RM 2
              ||
Db          2 RM 3


RESULT 7
US-10-032-717-48
; Sequence 48, Application US/10032717
; Publication No. US20020151709A1
; GENERAL INFORMATION:
;  APPLICANT: Andre R. Abad
;  APPLICANT:  Nicholas B. Duck
;  APPLICANT:  Xiang Feng
;  APPLICANT:  Ronald D. Flannagan
;  APPLICANT:  Theodore W. Kahn
;  APPLICANT:  Lynn E. Sims
;  TITLE OF INVENTION: Genes Encoding No. US20020151709A1el Proteins With
;  TITLE OF INVENTION:  Pesticidal Activity Against Coleopterans
;  FILE REFERENCE: 35718/237005
;  CURRENT APPLICATION NUMBER: US/10/032,717
;  CURRENT FILING DATE:  2001-10-23
;  PRIOR APPLICATION NUMBER: 60/242,838
;  PRIOR FILING DATE: 2000-10-24
;  NUMBER OF SEQ ID NOS: 48
;  SOFTWARE: FastSEQ for Windows Version 4.0
; SEQ ID NO 48
;   LENGTH: 4
;   TYPE: PRT
;   ORGANISM: Artificial Sequence
;   FEATURE:
;   OTHER INFORMATION: LRMS Insert
US-10-032-717-48

  Query Match             100.0%;  Score 10;  DB 4;  Length 4;
  Best Local Similarity   100.0%;  Pred. No. 1.7e+06;
  Matches    2;  Conservative    0;  Mismatches    0;  Indels    0;  Gaps    0;

Qy          1 RM 2
              ||
Db          2 RM 3


RESULT 8
US-10-197-000-3
; Sequence 3, Application US/10197000
; Publication No. US20030091582A1
; GENERAL INFORMATION:
;  APPLICANT: Humphreys, Robert E.
;  APPLICANT:  Adams, Sharlene
;  APPLICANT:  Xu, Minzhen
;  TITLE OF INVENTION: HYBRID PEPTIDES MODULATE THE IMMUNE RESPONSE
;  FILE REFERENCE: REH2007
;  CURRENT APPLICATION NUMBER: US/10/197,000
;  CURRENT FILING DATE:  2002-07-17
;  NUMBER OF SEQ ID NOS: 12
;  SOFTWARE: PatentIn Ver. 2.0
; SEQ ID NO 3
;   LENGTH: 4
;   TYPE: PRT
;   ORGANISM: Artificial Sequence
;   FEATURE:
;   OTHER INFORMATION: Description of Artificial Sequence: Modified mouse
;   OTHER INFORMATION: Ii key peptide
US-10-197-000-3

  Query Match             100.0%;  Score 10;  DB 4;  Length 4;
  Best Local Similarity   100.0%;  Pred. No. 1.7e+06;
  Matches    2;  Conservative    0;  Mismatches    0;  Indels    0;  Gaps    0;

Qy          1 RM 2
              ||
Db          2 RM 3


RESULT 9
US-10-163-198-86
; Sequence 86, Application US/10163198
; Publication No. US20030126645A1
; GENERAL INFORMATION:
;  APPLICANT: Rebecca E. Cahoon
;  APPLICANT:  Elmer P. Heppard
;  APPLICANT:  No. US20030126645A1uhiro Nagasawa
;  APPLICANT:  Hajime Sakai
;  TITLE OF INVENTION: Alteration Of Embryo/Endosperm Size During Seed Development
;  FILE REFERENCE: BB1487 US NA
;  CURRENT APPLICATION NUMBER: US/10/163,198
;  CURRENT FILING DATE:  2002-06-05
;  PRIOR APPLICATION NUMBER: 60/295,921
;  PRIOR FILING DATE: 2001-06-05
;  PRIOR APPLICATION NUMBER: 60/334,317
;  PRIOR FILING DATE: 2001-11-28
;  NUMBER OF SEQ ID NOS: 111
;  SOFTWARE: Microsoft Office 97
; SEQ ID NO 86
;   LENGTH: 4
;   TYPE: PRT
;   ORGANISM: conserved sequence motif
US-10-163-198-86
```

Query Match 100.0%; Score 10; DB 4; Length 4;
Best Local Similarity 100.0%; Pred. No. 1.7e+06;
Matches 2; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

Qy 1 RM 2
||
Db 3 RM 4

RESULT 10
US-10-385-317-2
; Sequence 2, Application US/10385317
; Publication No. US20030161857A1
; GENERAL INFORMATION:
; APPLICANT: Nachman, Ronald J
; APPLICANT: Garside, Christopher S
; APPLICANT: Tobe, Stephen S
; TITLE OF INVENTION: Mimetic Insect Allatostatin Analogs for Insect Control
; FILE REFERENCE: P.C. 0047.99-Ronald J. Nachman et al.
; CURRENT APPLICATION NUMBER: US/10/385,317
; CURRENT FILING DATE: 2003-03-10
; NUMBER OF SEQ ID NOS: 14
; SOFTWARE: PatentIn Ver. 2.1
; SEQ ID NO 2
; LENGTH: 4
; TYPE: PRT
; ORGANISM: Diploptera punctata
US-10-385-317-2

Query Match 100.0%; Score 10; DB 4; Length 4;
Best Local Similarity 100.0%; Pred. No. 1.7e+06;
Matches 2; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

Qy 1 RM 2
||
Db 3 RM 4

RESULT 11
US-10-414-637-48
; Sequence 48, Application US/10414637
; Publication No. US20030177528A1
; GENERAL INFORMATION:
; APPLICANT: Andre R. Abad
; APPLICANT: Nicholas B. Duck
; APPLICANT: Xiang Feng
; APPLICANT: Ronald D. Flannagan
; APPLICANT: Theodore W. Kahn
; APPLICANT: Lynn E. Sims
; TITLE OF INVENTION: Genes Encoding No. US20030177528A1el Proteins With
; TITLE OF INVENTION: Pesticidal Activity Against Coleopterans
; FILE REFERENCE: 35718/237005
; CURRENT APPLICATION NUMBER: US/10/414,637
; CURRENT FILING DATE: 2003-04-16
; PRIOR APPLICATION NUMBER: US/10/032,717
; PRIOR FILING DATE: 2001-10-23
; PRIOR APPLICATION NUMBER: 60/242,838
; PRIOR FILING DATE: 2000-10-24
; NUMBER OF SEQ ID NOS: 48
; SOFTWARE: FastSEQ for Windows Version 4.0
; SEQ ID NO 48
; LENGTH: 4
; TYPE: PRT
; ORGANISM: Artificial Sequence
; FEATURE:
; OTHER INFORMATION: LRMS Insert
US-10-414-637-48

Query Match 100.0%; Score 10; DB 4; Length 4;
Best Local Similarity 100.0%; Pred. No. 1.7e+06;
Matches 2; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

Qy 1 RM 2
||
Db 2 RM 3

RESULT 12
US-10-245-871-3
; Sequence 3, Application US/10245871
; Publication No. US20030235594A1
; GENERAL INFORMATION:
; APPLICANT: HUMPHREYS, ROBERT
; APPLICANT: XU, MINZHEN
; TITLE OF INVENTION: Ii-KEY/ANTIGENIC EPITOPE HYBRID PEPTIDE VACCINES
; FILE REFERENCE: REH-2013
; CURRENT APPLICATION NUMBER: US/10/245,871
; CURRENT FILING DATE: 2003-01-09
; PRIOR APPLICATION NUMBER: 10/197,000
; PRIOR FILING DATE: 2002-07-17
; PRIOR APPLICATION NUMBER: 09/396,813
; PRIOR FILING DATE: 1999-09-14
; NUMBER OF SEQ ID NOS: 905
; SOFTWARE: PatentIn Ver. 2.1
; SEQ ID NO 3
; LENGTH: 4
; TYPE: PRT
; ORGANISM: Artificial Sequence
; FEATURE:
; OTHER INFORMATION: Description of Artificial Sequence: Synthetic
; OTHER INFORMATION: Ii-key peptide
US-10-245-871-3

Query Match 100.0%; Score 10; DB 4; Length 4;
Best Local Similarity 100.0%; Pred. No. 1.7e+06;
Matches 2; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

Qy 1 RM 2
||
Db 2 RM 3

RESULT 13
US-10-245-871-8
; Sequence 8, Application US/10245871
; Publication No. US20030235594A1
; GENERAL INFORMATION:
; APPLICANT: HUMPHREYS, ROBERT
; APPLICANT: XU, MINZHEN
; TITLE OF INVENTION: Ii-KEY/ANTIGENIC EPITOPE HYBRID PEPTIDE VACCINES
; FILE REFERENCE: REH-2013
; CURRENT APPLICATION NUMBER: US/10/245,871
; CURRENT FILING DATE: 2003-01-09
; PRIOR APPLICATION NUMBER: 10/197,000
; PRIOR FILING DATE: 2002-07-17
; PRIOR APPLICATION NUMBER: 09/396,813
; PRIOR FILING DATE: 1999-09-14
; NUMBER OF SEQ ID NOS: 905
; SOFTWARE: PatentIn Ver. 2.1
; SEQ ID NO 8
; LENGTH: 4
; TYPE: PRT
; ORGANISM: Artificial Sequence
; FEATURE:
; OTHER INFORMATION: Description of Artificial Sequence: Synthetic
; OTHER INFORMATION: Ii-key peptide
US-10-245-871-8

Query Match 100.0%; Score 10; DB 4; Length 4;
Best Local Similarity 100.0%; Pred. No. 1.7e+06;
Matches 2; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

Qy 1 RM 2
||
Db 2 RM 3

```
RESULT 14
US-10-245-871-801
; Sequence 801, Application US/10245871
; Publication No. US20030235594A1
; GENERAL INFORMATION:
;  APPLICANT: HUMPHREYS, ROBERT
;  APPLICANT:  XU, MINZHEN
;  TITLE OF INVENTION: Ii-KEY/ANTIGENIC EPITOPE HYBRID PEPTIDE VACCINES
;  FILE REFERENCE: REH-2013
;  CURRENT APPLICATION NUMBER: US/10/245,871
;  CURRENT FILING DATE:  2003-01-09
;  PRIOR APPLICATION NUMBER: 10/197,000
;  PRIOR FILING DATE: 2002-07-17
;  PRIOR APPLICATION NUMBER: 09/396,813
;  PRIOR FILING DATE: 1999-09-14
;  NUMBER OF SEQ ID NOS: 905
;  SOFTWARE: PatentIn Ver. 2.1
; SEQ ID NO 801
;   LENGTH: 4
;   TYPE: PRT
;   ORGANISM: Artificial Sequence
;   FEATURE:
;   OTHER INFORMATION: Description of Artificial Sequence: Synthetic
;   OTHER INFORMATION: Ii-key/Ebola virus VP24 MHC Class II epitope
;   OTHER INFORMATION: hybrid peptide
;   FEATURE:
;   OTHER INFORMATION: C-term amidated
US-10-245-871-801

  Query Match             100.0%;  Score 10;  DB 4;  Length 4;
  Best Local Similarity   100.0%;  Pred. No. 1.7e+06;
  Matches    2;  Conservative    0;  Mismatches    0;  Indels    0;  Gaps    0;

Qy          1 RM 2
              ||
Db          2 RM 3


RESULT 15
US-10-659-509-2
; Sequence 2, Application US/10659509
; Publication No. US20040039159A1
; GENERAL INFORMATION:
;  APPLICANT: Nachman, Ronald J
;  APPLICANT:  Garside, Christopher S
;  APPLICANT:  Tobe, Stephen S
;  TITLE OF INVENTION: Mimetic Insect Allatostatin Analogs for Insect Control
;  FILE REFERENCE: P.C. 0047.99-Ronald J. Nachman et al.
;  CURRENT APPLICATION NUMBER: US/10/659,509
;  CURRENT FILING DATE:  2003-09-10
;  NUMBER OF SEQ ID NOS: 14
;  SOFTWARE: PatentIn Ver. 2.1
; SEQ ID NO 2
;   LENGTH: 4
;   TYPE: PRT
;   ORGANISM: Diploptera punctata
US-10-659-509-2

  Query Match             100.0%;  Score 10;  DB 4;  Length 4;
  Best Local Similarity   100.0%;  Pred. No. 1.7e+06;
  Matches    2;  Conservative    0;  Mismatches    0;  Indels    0;  Gaps    0;

Qy          1 RM 2
              ||
Db          3 RM 4


Search completed: February 28, 2006, 09:02:03
Job time : 1.58312 secs
```

This Page Blank (uspto)

GenCore version 5.1.7

OM protein - protein search, using sw model

Run on: February 28, 2006, 08:56:26 ; Search time 0.053467 Seconds
(without alignments)
556.876 Million cell updates/sec

Title: US-10-717-665A-44_COPY_448_449
Perfect score: 10
Sequence: 1 RM 2

Scoring table: BLOSUM62
Gapop 10.0 , Gapext 0.5

Searched: 117670 seqs, 14887254 residues

Total number of hits satisfying chosen parameters: 117670

Minimum DB seq length: 0
Maximum DB seq length: 2000000000

Post-processing: Minimum Match 0%
Maximum Match 100%
Listing first 45 summaries

Database : Published_Applications_AA_New:*
1: /cgn2_6/ptodata/2/pubpaa/US08_NEW_PUB.pep:*
2: /cgn2_6/ptodata/2/pubpaa/US06_NEW_PUB.pep:*
3: /cgn2_6/ptodata/2/pubpaa/US07_NEW_PUB.pep:*
4: /cgn2_6/ptodata/2/pubpaa/PCT_NEW_PUB.pep:*
5: /cgn2_6/ptodata/2/pubpaa/US09_NEW_PUB.pep:*
6: /cgn2_6/ptodata/2/pubpaa/US10_NEW_PUB.pep:*
7: /cgn2_6/ptodata/2/pubpaa/US11_NEW_PUB.pep:*
8: /cgn2_6/ptodata/2/pubpaa/US60_NEW_PUB.pep:*

Pred. No. is the number of results predicted by chance to have a score greater than or equal to the score of the result being printed, and is derived by analysis of the total score distribution.

SUMMARIES

Result No.	Score	% Query Match	Length	DB	ID	Description
1	10	100.0	4	6	US-10-895-064-1660	Sequence 1660, Ap
2	10	100.0	4	7	US-11-033-039-3	Sequence 3, Appli
3	10	100.0	4	7	US-11-033-039-8	Sequence 8, Appli
4	10	100.0	4	7	US-11-033-039-801	Sequence 801, App
5	10	100.0	4	7	US-11-129-741-1660	Sequence 1660, Ap
6	10	100.0	5	6	US-10-895-064-2685	Sequence 2685, Ap
7	10	100.0	5	7	US-11-062-186-20	Sequence 20, Appl
8	10	100.0	5	7	US-11-191-574-9	Sequence 9, Appli
9	10	100.0	5	7	US-11-033-039-9	Sequence 9, Appli
10	10	100.0	5	7	US-11-145-861-365	Sequence 365, App
11	10	100.0	5	7	US-11-129-741-2685	Sequence 2685, Ap
12	10	100.0	5	7	US-11-249-847-51	Sequence 51, Appl
13	10	100.0	5	7	US-11-249-847-486	Sequence 486, App
14	10	100.0	5	7	US-11-249-847-491	Sequence 491, App
15	10	100.0	5	7	US-11-249-847-493	Sequence 493, App
16	10	100.0	5	7	US-11-249-847-499	Sequence 499, App
17	10	100.0	5	7	US-11-249-847-528	Sequence 528, App
18	10	100.0	5	7	US-11-249-847-548	Sequence 548, App
19	10	100.0	5	7	US-11-249-847-561	Sequence 561, App
20	10	100.0	5	7	US-11-249-847-580	Sequence 580, App
21	10	100.0	7	6	US-10-467-657-9051	Sequence 9051, Ap
22	10	100.0	7	7	US-11-069-858-5	Sequence 5, Appli
23	10	100.0	7	7	US-11-069-858-6	Sequence 6, Appli
24	10	100.0	7	7	US-11-054-515-3182	Sequence 3182, Ap
25	10	100.0	7	7	US-11-033-039-4	Sequence 4, Appli
26	10	100.0	7	7	US-11-244-209-20	Sequence 20, Appl
27	10	100.0	7	7	US-11-247-423-459	Sequence 459, App
28	10	100.0	8	6	US-10-980-346B-43	Sequence 43, Appl
29	10	100.0	8	7	US-11-045-024-1087	Sequence 1087, Ap
30	10	100.0	8	7	US-11-045-024-1088	Sequence 1088, Ap
31	10	100.0	8	7	US-11-045-024-1089	Sequence 1089, Ap
32	10	100.0	8	7	US-11-045-024-1090	Sequence 1090, Ap
33	10	100.0	8	7	US-11-045-024-1091	Sequence 1091, Ap
34	10	100.0	8	7	US-11-045-024-1679	Sequence 1679, Ap
35	10	100.0	8	7	US-11-045-024-2508	Sequence 2508, Ap
36	10	100.0	8	7	US-11-045-024-2509	Sequence 2509, Ap
37	10	100.0	8	7	US-11-045-024-2510	Sequence 2510, Ap
38	10	100.0	8	7	US-11-045-024-3503	Sequence 3503, Ap
39	10	100.0	8	7	US-11-045-024-3504	Sequence 3504, Ap
40	10	100.0	8	7	US-11-045-024-3505	Sequence 3505, Ap
41	10	100.0	8	7	US-11-045-024-3506	Sequence 3506, Ap
42	10	100.0	8	7	US-11-045-024-4032	Sequence 4032, Ap
43	10	100.0	8	7	US-11-045-024-4033	Sequence 4033, Ap
44	10	100.0	8	7	US-11-045-024-4034	Sequence 4034, Ap
45	10	100.0	8	7	US-11-045-024-4369	Sequence 4369, Ap

ALIGNMENTS

RESULT 1
US-10-895-064-1660
; Sequence 1660, Application US/10895064
; Publication No. US20060018923A1
; GENERAL INFORMATION:
; APPLICANT: PEIRIS, JOSEPH S.M.
; APPLICANT: YUEN, KWOK YUNG
; APPLICANT: POON, LIT MAN
; APPLICANT: GUAN, YI
; APPLICANT: CHAN, KWOK HUNG
; APPLICANT: NICHOLLS, JOHN M.
; APPLICANT: LEUNG, FREDERICK C.
; TITLE OF INVENTION: A NOVEL HUMAN VIRUS CAUSING RESPIRATORY TRACT INFECTION AND USES
; TITLE OF INVENTION: THEREOF
; FILE REFERENCE: V0690.0031
; CURRENT APPLICATION NUMBER: US/10/895,064
; CURRENT FILING DATE: 2004-07-21
; NUMBER OF SEQ ID NOS: 2918
; SOFTWARE: PatentIn version 3.2
; SEQ ID NO 1660
; LENGTH: 4
; TYPE: PRT
; ORGANISM: Corononavirus-HKU1
US-10-895-064-1660

Query Match 100.0%; Score 10; DB 6; Length 4;
Best Local Similarity 100.0%; Pred. No. 8.2e+04;
Matches 2; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

Qy 1 RM 2
||
Db 1 RM 2

RESULT 2
US-11-033-039-3
; Sequence 3, Application US/11033039
; Publication No. US20060002947A1
; GENERAL INFORMATION:
; APPLICANT: HUMPHREYS, ROBERT
; APPLICANT: XU, MINZHEN
; TITLE OF INVENTION: LI-KEY/ANTIGENIC EPITOPE HYBRID PEPTIDE VACCINES
; FILE REFERENCE: REH-2017US01
; CURRENT APPLICATION NUMBER: US/11/033,039
; CURRENT FILING DATE: 2005-01-11
; PRIOR APPLICATION NUMBER: 10/245,871
; PRIOR FILING DATE: 2002-09-17
; PRIOR APPLICATION NUMBER: 10/197,000

; PRIOR FILING DATE: 2002-07-17
; PRIOR APPLICATION NUMBER: 09/396,813
; PRIOR FILING DATE: 1999-09-14
; NUMBER OF SEQ ID NOS: 1452
; SOFTWARE: PatentIn version 3.3
; SEQ ID NO 3
; LENGTH: 4
; TYPE: PRT
; ORGANISM: Artificial Sequence
; FEATURE:
; OTHER INFORMATION: Description of Artificial Sequence: Modified
; OTHER INFORMATION: mammalian li-key peptide
US-11-033-039-3

Query Match 100.0%; Score 10; DB 7; Length 4;
Best Local Similarity 100.0%; Pred. No. 8.2e+04;
Matches 2; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

Qy 1 RM 2
||
Db 2 RM 3

RESULT 3
US-11-033-039-8
; Sequence 8, Application US/11033039
; Publication No. US20060002947A1
; GENERAL INFORMATION:
; APPLICANT: HUMPHREYS, ROBERT
; APPLICANT: XU, MINZHEN
; TITLE OF INVENTION: LI-KEY/ANTIGENIC EPITOPE HYBRID PEPTIDE VACCINES
; FILE REFERENCE: REH-2017US01
; CURRENT APPLICATION NUMBER: US/11/033,039
; CURRENT FILING DATE: 2005-01-11
; PRIOR APPLICATION NUMBER: 10/245,871
; PRIOR FILING DATE: 2002-09-17
; PRIOR APPLICATION NUMBER: 10/197,000
; PRIOR FILING DATE: 2002-07-17
; PRIOR APPLICATION NUMBER: 09/396,813
; PRIOR FILING DATE: 1999-09-14
; NUMBER OF SEQ ID NOS: 1452
; SOFTWARE: PatentIn version 3.3
; SEQ ID NO 8
; LENGTH: 4
; TYPE: PRT
; ORGANISM: Artificial Sequence
; FEATURE:
; OTHER INFORMATION: Description of Artificial Sequence: Modified
; OTHER INFORMATION: mammalian li-key peptide
US-11-033-039-8

Query Match 100.0%; Score 10; DB 7; Length 4;
Best Local Similarity 100.0%; Pred. No. 8.2e+04;
Matches 2; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

Qy 1 RM 2
||
Db 2 RM 3

RESULT 4
US-11-033-039-801
; Sequence 801, Application US/11033039
; Publication No. US20060002947A1
; GENERAL INFORMATION:
; APPLICANT: HUMPHREYS, ROBERT
; APPLICANT: XU, MINZHEN
; TITLE OF INVENTION: LI-KEY/ANTIGENIC EPITOPE HYBRID PEPTIDE VACCINES
; FILE REFERENCE: REH-2017US01
; CURRENT APPLICATION NUMBER: US/11/033,039
; CURRENT FILING DATE: 2005-01-11
; PRIOR APPLICATION NUMBER: 10/245,871
; PRIOR FILING DATE: 2002-09-17
; PRIOR APPLICATION NUMBER: 10/197,000
; PRIOR FILING DATE: 2002-07-17
; PRIOR APPLICATION NUMBER: 09/396,813
; PRIOR FILING DATE: 1999-09-14
; NUMBER OF SEQ ID NOS: 1452
; SOFTWARE: PatentIn version 3.3
; SEQ ID NO 801
; LENGTH: 4
; TYPE: PRT
; ORGANISM: Artificial Sequence
; FEATURE:
; OTHER INFORMATION: Description of Artificial Sequence: Synthetic
; OTHER INFORMATION: hybrid peptide
US-11-033-039-801

Query Match 100.0%; Score 10; DB 7; Length 4;
Best Local Similarity 100.0%; Pred. No. 8.2e+04;
Matches 2; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

Qy 1 RM 2
||
Db 2 RM 3

RESULT 5
US-11-129-741-1660
; Sequence 1660, Application US/11129741
; Publication No. US20060034853A1
; GENERAL INFORMATION:
; APPLICANT: YUEN, KWOK YUNG
; APPLICANT: WOO, CHIU YAT PATRICK
; APPLICANT: LAU, KAR PUI SUSANNA
; APPLICANT: CHAN, KWOK HUNG
; APPLICANT: POON, LIT MAN
; APPLICANT: PEIRIS, JOSEPH S.M.
; APPLICANT: GUAN, YI
; TITLE OF INVENTION: A NOVEL HUMAN VIRUS CAUSING RESPIRATORY TRACT
; TITLE OF INVENTION: INFECTION AND USES THEREOF
; FILE REFERENCE: V0690.0044
; CURRENT APPLICATION NUMBER: US/11/129,741
; CURRENT FILING DATE: 2005-05-16
; PRIOR APPLICATION NUMBER: 10/895,064
; PRIOR FILING DATE: 2004-07-21
; NUMBER OF SEQ ID NOS: 4257
; SOFTWARE: PatentIn version 3.3
; SEQ ID NO 1660
; LENGTH: 4
; TYPE: PRT
; ORGANISM: Corononavirus-HKU1
US-11-129-741-1660

Query Match 100.0%; Score 10; DB 7; Length 4;
Best Local Similarity 100.0%; Pred. No. 8.2e+04;
Matches 2; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

Qy 1 RM 2
||
Db 1 RM 2

RESULT 6
US-10-895-064-2685
; Sequence 2685, Application US/10895064
; Publication No. US20060018923A1
; GENERAL INFORMATION:
; APPLICANT: PEIRIS, JOSEPH S.M.
; APPLICANT: YUEN, KWOK YUNG
; APPLICANT: POON, LIT MAN
; APPLICANT: GUAN, YI
; APPLICANT: CHAN, KWOK HUNG
; APPLICANT: NICHOLLS, JOHN M.
; APPLICANT: LEUNG, FREDERICK C.
; TITLE OF INVENTION: A NOVEL HUMAN VIRUS CAUSING RESPIRATORY TRACT INFECTION AND USES

; TITLE OF INVENTION: THEREOF
; FILE REFERENCE: V0690.0031
; CURRENT APPLICATION NUMBER: US/10/895,064
; CURRENT FILING DATE: 2004-07-21
; NUMBER OF SEQ ID NOS: 2918
; SOFTWARE: PatentIn version 3.2
; SEQ ID NO 2685
; LENGTH: 5
; TYPE: PRT
; ORGANISM: Coronavirus-HKU1
US-10-895-064-2685

Query Match 100.0%; Score 10; DB 6; Length 5;
Best Local Similarity 100.0%; Pred. No. 8.2e+04;
Matches 2; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

Qy 1 RM 2
||
Db 4 RM 5

RESULT 7
US-11-062-186-20
; Sequence 20, Application US/11062186
; Publication No. US20050272097A1
; GENERAL INFORMATION:
; APPLICANT: CALENOFF, EMANUEL
; TITLE OF INVENTION: METHODS AND COMPOSITIONS FOR DETECTING AND TREATING
; TITLE OF INVENTION: AUTOIMMUNE DISEASES
; FILE REFERENCE: 21417-98470
; CURRENT APPLICATION NUMBER: US/11/062,186
; CURRENT FILING DATE: 2005-02-18
; PRIOR APPLICATION NUMBER: 60/546,062
; PRIOR FILING DATE: 2004-02-18
; PRIOR APPLICATION NUMBER: 60/545,980
; PRIOR FILING DATE: 2004-02-18
; NUMBER OF SEQ ID NOS: 172
; SOFTWARE: PatentIn version 3.3
; SEQ ID NO 20
; LENGTH: 5
; TYPE: PRT
; ORGANISM: Homo sapiens
US-11-062-186-20

Query Match 100.0%; Score 10; DB 7; Length 5;
Best Local Similarity 100.0%; Pred. No. 8.2e+04;
Matches 2; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

Qy 1 RM 2
||
Db 4 RM 5

RESULT 8
US-11-191-574-9
; Sequence 9, Application US/11191574
; Publication No. US20050272125A1
; GENERAL INFORMATION:
; APPLICANT: Hoeg-Jensen, Thomas
; APPLICANT: Egel-Mitani, Michi
; APPLICANT: Balschmidt, Per
; APPLICANT: Markussen, Jan
; APPLICANT: Diers, Ivan
; TITLE OF INVENTION: Method for Making Acylated Polypeptides
; FILE REFERENCE: 6305.200-US
; CURRENT APPLICATION NUMBER: US/11/191,574
; CURRENT FILING DATE: 2005-07-28
; PRIOR APPLICATION NUMBER: US/10/205,270
; PRIOR FILING DATE: 2002-07-24
; PRIOR APPLICATION NUMBER: PA 2001 01140
; PRIOR FILING DATE: 2001-07-24
; PRIOR APPLICATION NUMBER: 60/310,952
; PRIOR FILING DATE: 2001-08-08
; NUMBER OF SEQ ID NOS: 16
; SOFTWARE: FastSEQ for Windows Version 4.0
; SEQ ID NO 9
; LENGTH: 5
; TYPE: PRT
; ORGANISM: Artificial Sequence
; FEATURE:
; OTHER INFORMATION: Synthetic
US-11-191-574-9

Query Match 100.0%; Score 10; DB 7; Length 5;
Best Local Similarity 100.0%; Pred. No. 8.2e+04;
Matches 2; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

Qy 1 RM 2
||
Db 4 RM 5

RESULT 9
US-11-033-039-9
; Sequence 9, Application US/11033039
; Publication No. US20060002947A1
; GENERAL INFORMATION:
; APPLICANT: HUMPHREYS, ROBERT
; APPLICANT: XU, MINZHEN
; TITLE OF INVENTION: LI-KEY/ANTIGENIC EPITOPE HYBRID PEPTIDE VACCINES
; FILE REFERENCE: REH-2017US01
; CURRENT APPLICATION NUMBER: US/11/033,039
; CURRENT FILING DATE: 2005-01-11
; PRIOR APPLICATION NUMBER: 10/245,871
; PRIOR FILING DATE: 2002-09-17
; PRIOR APPLICATION NUMBER: 10/197,000
; PRIOR FILING DATE: 2002-07-17
; PRIOR APPLICATION NUMBER: 09/396,813
; PRIOR FILING DATE: 1999-09-14
; NUMBER OF SEQ ID NOS: 1452
; SOFTWARE: PatentIn version 3.3
; SEQ ID NO 9
; LENGTH: 5
; TYPE: PRT
; ORGANISM: Artificial Sequence
; FEATURE:
; OTHER INFORMATION: Description of Artificial Sequence: Modified
; OTHER INFORMATION: mammalian li-key peptide
; FEATURE:
; NAME/KEY: MOD_RES
; LOCATION: (5)..(5)
; OTHER INFORMATION: Ava
US-11-033-039-9

Query Match 100.0%; Score 10; DB 7; Length 5;
Best Local Similarity 100.0%; Pred. No. 8.2e+04;
Matches 2; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

Qy 1 RM 2
||
Db 2 RM 3

RESULT 10
US-11-145-861-365
; Sequence 365, Application US/11145861
; Publication No. US20060014138A1
; GENERAL INFORMATION:
; APPLICANT: Chinnaiyan, Arul
; APPLICANT: Wang, Xiaoju
; TITLE OF INVENTION: Phage Microarray Profiling of the Humoral Response to Disease
; FILE REFERENCE: UM-09899
; CURRENT APPLICATION NUMBER: US/11/145,861
; CURRENT FILING DATE: 2005-06-06
; NUMBER OF SEQ ID NOS: 464
; SOFTWARE: PatentIn version 3.3

; SEQ ID NO 365
; LENGTH: 5
; TYPE: PRT
; ORGANISM: Homo sapiens
US-11-145-861-365

Query Match 100.0%; Score 10; DB 7; Length 5;
Best Local Similarity 100.0%; Pred. No. 8.2e+04;
Matches 2; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

Qy 1 RM 2
||
Db 3 RM 4

RESULT 11
US-11-129-741-2685
; Sequence 2685, Application US/11129741
; Publication No. US20060034853A1
; GENERAL INFORMATION:
; APPLICANT: YUEN, KWOK YUNG
; APPLICANT: WOO, CHIU YAT PATRICK
; APPLICANT: LAU, KAR PUI SUSANNA
; APPLICANT: CHAN, KWOK HUNG
; APPLICANT: POON, LIT MAN
; APPLICANT: PEIRIS, JOSEPH S.M.
; APPLICANT: GUAN, YI
; TITLE OF INVENTION: A NOVEL HUMAN VIRUS CAUSING RESPIRATORY TRACT
; TITLE OF INVENTION: INFECTION AND USES THEREOF
; FILE REFERENCE: V0690.0044
; CURRENT APPLICATION NUMBER: US/11/129,741
; CURRENT FILING DATE: 2005-05-16
; PRIOR APPLICATION NUMBER: 10/895,064
; PRIOR FILING DATE: 2004-07-21
; NUMBER OF SEQ ID NOS: 4257
; SOFTWARE: PatentIn version 3.3
; SEQ ID NO 2685
; LENGTH: 5
; TYPE: PRT
; ORGANISM: Corononavirus-HKU1
US-11-129-741-2685

Query Match 100.0%; Score 10; DB 7; Length 5;
Best Local Similarity 100.0%; Pred. No. 8.2e+04;
Matches 2; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

Qy 1 RM 2
||
Db 4 RM 5

RESULT 12
US-11-249-847-51
; Sequence 51, Application US/11249847
; Publication No. US20060035270A1
; GENERAL INFORMATION:
; APPLICANT: Lee, Frank D.
; APPLICANT: Meng, Dr. Xun
; APPLICANT: Chan, John W.
; APPLICANT: Zhang, Shengsheng
; APPLICANT: Benkovic, Stephen J.
; TITLE OF INVENTION: UNIQUE RECOGNITION SEQUENCES AND METHODS OF USE THEREOF IN
; TITLE OF INVENTION: PROTEIN ANALYSIS
; FILE REFERENCE: EPTM-P05-001
; CURRENT APPLICATION NUMBER: US/11/249,847
; CURRENT FILING DATE: 2005-10-12
; PRIOR APPLICATION NUMBER: 10/436,549
; PRIOR FILING DATE: 2003-05-12
; PRIOR APPLICATION NUMBER: 60/379,626
; PRIOR FILING DATE: 2002-05-10
; PRIOR APPLICATION NUMBER: 60/393,137
; PRIOR FILING DATE: 2002-07-01
; PRIOR APPLICATION NUMBER: 60/393,197
; PRIOR FILING DATE: 2002-07-01
; PRIOR APPLICATION NUMBER: 60/393,211
; PRIOR FILING DATE: 2002-07-01
; PRIOR APPLICATION NUMBER: 60/393,223
; PRIOR FILING DATE: 2002-07-01
; PRIOR APPLICATION NUMBER: 60/393,233
; PRIOR FILING DATE: 2002-07-01
; PRIOR APPLICATION NUMBER: 60/393,235
; PRIOR FILING DATE: 2002-07-01
; PRIOR APPLICATION NUMBER: 60/393,280
; PRIOR FILING DATE: 2002-07-01
; PRIOR APPLICATION NUMBER: 60/430,948
; PRIOR FILING DATE: 2002-12-04
; Remaining Prior Application data removed - See File Wrapper or PALM.
; NUMBER OF SEQ ID NOS: 614
; SOFTWARE: PatentIn version 3.2
; SEQ ID NO 51
; LENGTH: 5
; TYPE: PRT
; ORGANISM: Homo sapiens
US-11-249-847-51

Query Match 100.0%; Score 10; DB 7; Length 5;
Best Local Similarity 100.0%; Pred. No. 8.2e+04;
Matches 2; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

Qy 1 RM 2
||
Db 4 RM 5

RESULT 13
US-11-249-847-486
; Sequence 486, Application US/11249847
; Publication No. US20060035270A1
; GENERAL INFORMATION:
; APPLICANT: Lee, Frank D.
; APPLICANT: Meng, Dr. Xun
; APPLICANT: Chan, John W.
; APPLICANT: Zhang, Shengsheng
; APPLICANT: Benkovic, Stephen J.
; TITLE OF INVENTION: UNIQUE RECOGNITION SEQUENCES AND METHODS OF USE THEREOF IN
; TITLE OF INVENTION: PROTEIN ANALYSIS
; FILE REFERENCE: EPTM-P05-001
; CURRENT APPLICATION NUMBER: US/11/249,847
; CURRENT FILING DATE: 2005-10-12
; PRIOR APPLICATION NUMBER: 10/436,549
; PRIOR FILING DATE: 2003-05-12
; PRIOR APPLICATION NUMBER: 60/379,626
; PRIOR FILING DATE: 2002-05-10
; PRIOR APPLICATION NUMBER: 60/393,137
; PRIOR FILING DATE: 2002-07-01
; PRIOR APPLICATION NUMBER: 60/393,197
; PRIOR FILING DATE: 2002-07-01
; PRIOR APPLICATION NUMBER: 60/393,211
; PRIOR FILING DATE: 2002-07-01
; PRIOR APPLICATION NUMBER: 60/393,223
; PRIOR FILING DATE: 2002-07-01
; PRIOR APPLICATION NUMBER: 60/393,233
; PRIOR FILING DATE: 2002-07-01
; PRIOR APPLICATION NUMBER: 60/393,235
; PRIOR FILING DATE: 2002-07-01
; PRIOR APPLICATION NUMBER: 60/393,280
; PRIOR FILING DATE: 2002-07-01
; PRIOR APPLICATION NUMBER: 60/430,948
; PRIOR FILING DATE: 2002-12-04
; Remaining Prior Application data removed - See File Wrapper or PALM.
; NUMBER OF SEQ ID NOS: 614
; SOFTWARE: PatentIn version 3.2
; SEQ ID NO 486
; LENGTH: 5
; TYPE: PRT
; ORGANISM: Homo sapiens

US-11-249-847-486

Query Match 100.0%; Score 10; DB 7; Length 5;
Best Local Similarity 100.0%; Pred. No. 8.2e+04;
Matches 2; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

```
Qy          1 RM 2
              ||
Db          4 RM 5
```

RESULT 14
US-11-249-847-491
; Sequence 491, Application US/11249847
; Publication No. US20060035270A1
; GENERAL INFORMATION:
; APPLICANT: Lee, Frank D.
; APPLICANT: Meng, Dr. Xun
; APPLICANT: Chan, John W.
; APPLICANT: Zhang, Shengsheng
; APPLICANT: Benkovic, Stephen J.
; TITLE OF INVENTION: UNIQUE RECOGNITION SEQUENCES AND METHODS OF USE THEREOF IN
; TITLE OF INVENTION: PROTEIN ANALYSIS
; FILE REFERENCE: EPTM-P05-001
; CURRENT APPLICATION NUMBER: US/11/249,847
; CURRENT FILING DATE: 2005-10-12
; PRIOR APPLICATION NUMBER: 10/436,549
; PRIOR FILING DATE: 2003-05-12
; PRIOR APPLICATION NUMBER: 60/379,626
; PRIOR FILING DATE: 2002-05-10
; PRIOR APPLICATION NUMBER: 60/393,137
; PRIOR FILING DATE: 2002-07-01
; PRIOR APPLICATION NUMBER: 60/393,197
; PRIOR FILING DATE: 2002-07-01
; PRIOR APPLICATION NUMBER: 60/393,211
; PRIOR FILING DATE: 2002-07-01
; PRIOR APPLICATION NUMBER: 60/393,223
; PRIOR FILING DATE: 2002-07-01
; PRIOR APPLICATION NUMBER: 60/393,233
; PRIOR FILING DATE: 2002-07-01
; PRIOR APPLICATION NUMBER: 60/393,235
; PRIOR FILING DATE: 2002-07-01
; PRIOR APPLICATION NUMBER: 60/393,280
; PRIOR FILING DATE: 2002-07-01
; PRIOR APPLICATION NUMBER: 60/430,948
; PRIOR FILING DATE: 2002-12-04
; Remaining Prior Application data removed - See File Wrapper or PALM.
; NUMBER OF SEQ ID NOS: 614
; SOFTWARE: PatentIn version 3.2
; SEQ ID NO 491
; LENGTH: 5
; TYPE: PRT
; ORGANISM: Homo sapiens
US-11-249-847-491

Query Match 100.0%; Score 10; DB 7; Length 5;
Best Local Similarity 100.0%; Pred. No. 8.2e+04;
Matches 2; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

```
Qy          1 RM 2
              ||
Db          4 RM 5
```

RESULT 15
US-11-249-847-493
; Sequence 493, Application US/11249847
; Publication No. US20060035270A1
; GENERAL INFORMATION:
; APPLICANT: Lee, Frank D.
; APPLICANT: Meng, Dr. Xun
; APPLICANT: Chan, John W.
; APPLICANT: Zhang, Shengsheng
; APPLICANT: Benkovic, Stephen J.
; TITLE OF INVENTION: UNIQUE RECOGNITION SEQUENCES AND METHODS OF USE THEREOF IN
; TITLE OF INVENTION: PROTEIN ANALYSIS
; FILE REFERENCE: EPTM-P05-001
; CURRENT APPLICATION NUMBER: US/11/249,847
; CURRENT FILING DATE: 2005-10-12
; PRIOR APPLICATION NUMBER: 10/436,549
; PRIOR FILING DATE: 2003-05-12
; PRIOR APPLICATION NUMBER: 60/379,626
; PRIOR FILING DATE: 2002-05-10
; PRIOR APPLICATION NUMBER: 60/393,137
; PRIOR FILING DATE: 2002-07-01
; PRIOR APPLICATION NUMBER: 60/393,197
; PRIOR FILING DATE: 2002-07-01
; PRIOR APPLICATION NUMBER: 60/393,211
; PRIOR FILING DATE: 2002-07-01
; PRIOR APPLICATION NUMBER: 60/393,223
; PRIOR FILING DATE: 2002-07-01
; PRIOR APPLICATION NUMBER: 60/393,233
; PRIOR FILING DATE: 2002-07-01
; PRIOR APPLICATION NUMBER: 60/393,235
; PRIOR FILING DATE: 2002-07-01
; PRIOR APPLICATION NUMBER: 60/393,280
; PRIOR FILING DATE: 2002-07-01
; PRIOR APPLICATION NUMBER: 60/430,948
; PRIOR FILING DATE: 2002-12-04
; Remaining Prior Application data removed - See File Wrapper or PALM.
; NUMBER OF SEQ ID NOS: 614
; SOFTWARE: PatentIn version 3.2
; SEQ ID NO 493
; LENGTH: 5
; TYPE: PRT
; ORGANISM: Homo sapiens
US-11-249-847-493

Query Match 100.0%; Score 10; DB 7; Length 5;
Best Local Similarity 100.0%; Pred. No. 8.2e+04;
Matches 2; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

```
Qy          1 RM 2
              ||
Db          2 RM 3
```

Search completed: February 28, 2006, 09:02:40
Job time : 0.053467 secs

GenCore version 5.1.7

OM protein - protein search, using sw model

Run on: February 28, 2006, 08:45:35 ; Search time 0.121972 Seconds
(without alignments)
1577.691 Million cell updates/sec

Title: US-10-717-665A-44_COPY_448_449
Perfect score: 10
Sequence: 1 RM 2

Scoring table: BLOSUM62
Gapop 10.0 , Gapext 0.5

Searched: 283416 seqs, 96216763 residues

Total number of hits satisfying chosen parameters: 283416

Minimum DB seq length: 0
Maximum DB seq length: 2000000000

Post-processing: Minimum Match 0%
Maximum Match 100%
Listing first 45 summaries

Database : PIR_80:*
1: pir1:*
2: pir2:*
3: pir3:*
4: pir4:*

Pred. No. is the number of results predicted by chance to have a score greater than or equal to the score of the result being printed, and is derived by analysis of the total score distribution.

SUMMARIES

Result No.	Score	% Query Match	Length	DB	ID	Description
1	10	100.0	8	2	PT0298	Ig heavy chain CRD
2	10	100.0	9	2	PT0670	T-cell receptor be
3	10	100.0	10	2	PC4442	cytochrome c553 -
4	10	100.0	11	2	E60691	phycobilisome 8K l
5	10	100.0	12	2	A42324	cytochrome P450c27
6	10	100.0	12	2	T44420	hypothetical prote
7	10	100.0	14	2	PC1215	homeotic protein E
8	10	100.0	14	2	H83778	hypothetical prote
9	10	100.0	15	2	A56786	pimeloyl-CoA synth
10	10	100.0	15	2	PA0020	protein QA100028 -
11	10	100.0	15	2	A45096	thyrotropin-releas
12	10	100.0	16	2	S24667	protein-tyrosine k
13	10	100.0	16	2	S11805	heat shock protein
14	10	100.0	16	2	A41170	photosystem II 6.1
15	10	100.0	16	2	S78415	ribosomal protein
16	10	100.0	16	2	E37290	homeotic protein G
17	10	100.0	17	2	A34835	ribosomal protein
18	10	100.0	17	2	S77834	DNA-directed RNA p
19	10	100.0	18	2	B24735	glutathione transf
20	10	100.0	19	2	S32548	glutathione transf
21	10	100.0	19	2	S31613	beta-1,3-glucanase
22	10	100.0	20	2	S29100	glutathione transf
23	10	100.0	20	2	S30381	glutathione transf
24	10	100.0	20	2	E60894	crystallin - Pacif
25	10	100.0	20	2	A44773	pollen allergen I
26	10	100.0	20	2	PC4385	GroEL protein homo
27	10	100.0	20	2	PC4386	GroEL protein homo
28	10	100.0	20	2	E54226	light-harvesting p
29	10	100.0	20	2	A56899	serum heterodimer,
30	10	100.0	20	2	A55899	basement membrane
31	10	100.0	22	2	JC0009	phospholipase A2 (
32	10	100.0	22	2	B58793	relaxin chain A -
33	10	100.0	23	2	H24735	glutathione transf
34	10	100.0	23	2	I48936	cryptdin - mouse (
35	10	100.0	23	2	I49413	cryptdin - western
36	10	100.0	23	2	S72535	probable acr-2 reg
37	10	100.0	23	2	A41263	kinase-related tra
38	10	100.0	23	2	S43469	heterodisulfide re
39	10	100.0	24	2	S30373	glutathione transf
40	10	100.0	24	2	T42441	protein phosphatas
41	10	100.0	24	2	A33262	heparin-binding gr
42	10	100.0	24	2	H53578	brevinin-1Ea - edi
43	10	100.0	24	2	T07991	hypothetical prote
44	10	100.0	25	1	JC4278	ribosomal protein
45	10	100.0	25	2	A48401	ribosomal protein

ALIGNMENTS

RESULT 1
PT0298
Ig heavy chain CRD3 region (clone 5-103A) - human (fragment)
C;Species: Homo sapiens (man)
C;Date: 30-Sep-1993 #sequence_revision 30-Sep-1993 #text_change 16-Aug-1996
C;Accession: PT0298
R;Yamada, M.; Wasserman, R.; Reichard, B.A.; Shane, S.; Caton, A.J.; Rovera, G.
J. Exp. Med. 173, 395-407, 1991
A;Title: Preferential utilization of specific immunoglobulin heavy chain diversity and jc
A;Reference number: PT0222; MUID:91108337; PMID:1899102
A;Accession: PT0298
A;Molecule type: DNA
A;Residues: 1-8 <YAM>
A;Cross-references: UNIPARC:UPI000017C208
A;Experimental source: B lymphocyte
C;Keywords: heterotetramer; immunoglobulin

Query Match 100.0%; Score 10; DB 2; Length 8;
Best Local Similarity 100.0%; Pred. No. 2.8e+05;
Matches 2; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

Qy 1 RM 2
||
Db 1 RM 2

RESULT 2
PT0670
T-cell receptor beta chain V-D-J region (121-1BN) - mouse (fragment)
C;Species: Mus musculus (house mouse)
C;Date: 17-Jul-1992 #sequence_revision 17-Jul-1992 #text_change 30-May-1997
C;Accession: PT0670
R;Feeney, A.J.
J. Exp. Med. 174, 115-124, 1991
A;Title: Junctional sequences of fetal T cell receptor beta chains have few N regions.
A;Reference number: PT0509; MUID:91277601; PMID:1711558
A;Accession: PT0670
A;Status: translation not shown
A;Molecule type: mRNA
A;Residues: 1-9 <FEE>
A;Cross-references: UNIPARC:UPI000017C7EC
A;Experimental source: day 4 postnatal thymus, strain BALB/c
C;Keywords: T-cell receptor

Query Match 100.0%; Score 10; DB 2; Length 9;
Best Local Similarity 100.0%; Pred. No. 2.8e+05;
Matches 2; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

Qy 1 RM 2
||
Db 8 RM 9

RESULT 3
PC4442
cytochrome c553 - Desulfovibrio desulfuricans (fragment)
C;Species: Desulfovibrio desulfuricans
C;Date: 28-Mar-1998 #sequence_revision 28-Mar-1998 #text_change 09-Jul-2004
C;Accession: PC4442
R;Aubert, C.; Leroy, G.; Bianco, P.; Forest, E.; Bruschi, M.; Dolla, A.
Biochem. Biophys. Res. Commun. 242, 213-218, 1998
A;Title: Characterization of the cytochromes C from Desulfovibrio desulfuricans G201.
A;Reference number: PC4442; MUID:98102811; PMID:9439638
A;Accession: PC4442
A;Molecule type: protein
A;Residues: 1-10 <AUB>
A;Cross-references: UNIPROT:Q7M0M6; UNIPARC:UPI000017AB62
A;Experimental source: strain G201
C;Comment: This protein is involved in the formate reduction pathway.

Query Match 100.0%; Score 10; DB 2; Length 10;
Best Local Similarity 100.0%; Pred. No. 7.6e+02;
Matches 2; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

Qy 1 RM 2
||
Db 9 RM 10

RESULT 4
E60691
phycobilisome 8K linker protein - Synechococcus sp. (PCC 7002) (fragment)
C;Species: Synechococcus sp.
C;Date: 14-May-1993 #sequence_revision 14-May-1993 #text_change 07-May-1999
C;Accession: E60691
R;Bryant, D.A.; de Lorimier, R.; Guglielmi, G.; Stevens Jr., S.E.
Arch. Microbiol. 153, 550-560, 1990
A;Title: Structural and compositional analyses of the phycobilisomes of Synechococcus sp
s.
A;Reference number: A60691; MUID:90314662; PMID:2164365
A;Accession: E60691
A;Molecule type: protein
A;Residues: 1-11 <BRY>
A;Cross-references: UNIPARC:UPI000017ABC5
C;Comment: This protein, one of the eleven components detected in this species of the ph
C;Keywords: photosystem II

Query Match 100.0%; Score 10; DB 2; Length 11;
Best Local Similarity 100.0%; Pred. No. 8.3e+02;
Matches 2; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

Qy 1 RM 2
||
Db 2 RM 3

RESULT 5
A42324
cytochrome P450c27/25 - rat (fragment)
C;Species: Rattus norvegicus (Norway rat)
C;Date: 04-Mar-1993 #sequence_revision 18-Nov-1994 #text_change 08-Feb-1996
C;Accession: A42324
R;Shayiq, R.M.; Avadhani, N.G.
J. Biol. Chem. 267, 2421-2428, 1992
A;Title: Sequence complementarity between the 5'-terminal regions of mRNAs for rat mitoc
ap.
A;Reference number: A42324; MUID:92129322; PMID:1733943
A;Accession: A42324
A;Status: preliminary
A;Molecule type: mRNA
A;Residues: 1-12 <SHA>
A;Cross-references: UNIPARC:UPI000017C8FA
A;Note: sequence extracted from NCBI backbone (NCBIN:78404, NCBIP:78410)

Query Match 100.0%; Score 10; DB 2; Length 12;
Best Local Similarity 100.0%; Pred. No. 9.1e+02;
Matches 2; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

Qy 1 RM 2
||
Db 6 RM 7

RESULT 6
T44420
hypothetical protein [imported] - Bacillus stearothermophilus (fragment)
C;Species: Bacillus stearothermophilus
C;Date: 21-Jan-2000 #sequence_revision 21-Jan-2000 #text_change 09-Jul-2004
C;Accession: T44420
R;Vlaskova, H.; Krasny, L.; Fucik, V.; Jonak, J.
submitted to the EMBL Data Library, September 1997
A;Description: The pyrAb gene coding for the large subunit of carbamoylphosphate synthet
A;Reference number: Z22760
A;Accession: T44420
A;Status: preliminary; translated from GB/EMBL/DDBJ
A;Molecule type: DNA
A;Residues: 1-12 <VLA>
A;Cross-references: UNIPROT:O50303; UNIPARC:UPI00000B72C5; EMBL:AJ001805; PIDN:CAA05021.1
A;Experimental source: strain CCM 2184
C;Genetics:
A;Note: ORF2

Query Match 100.0%; Score 10; DB 2; Length 12;
Best Local Similarity 100.0%; Pred. No. 9.1e+02;
Matches 2; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

Qy 1 RM 2
||
Db 6 RM 7

RESULT 7
PC1215
homeotic protein EgHbx5 - tapeworm (Echinococcus granulosus) (fragment)
C;Species: Echinococcus granulosus
C;Date: 30-Sep-1993 #sequence_revision 30-Sep-1993 #text_change 30-Sep-1993
C;Accession: PC1215
R;Oliver, G.; Vispo, M.; Mailhos, A.; Martinez, C.; Sosa-Pineda, B.; Fielitz, W.; Ehrlich
Gene 121, 337-342, 1992
A;Title: Homeoboxes in flatworms.
A;Reference number: JC1386; MUID:93077050; PMID:1359988
A;Accession: PC1215
A;Molecule type: DNA
A;Residues: 1-14 <OLI>
A;Cross-references: UNIPARC:UPI000017B68F; EMBL:X66821
C;Keywords: homeobox

Query Match 100.0%; Score 10; DB 2; Length 14;
Best Local Similarity 100.0%; Pred. No. 1e+03;
Matches 2; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

Qy 1 RM 2
||
Db 12 RM 13

RESULT 8
H83778
hypothetical protein BH1032 [imported] - Bacillus halodurans (strain C-125)
C;Species: Bacillus halodurans
C;Date: 01-Dec-2000 #sequence_revision 01-Dec-2000 #text_change 09-Jul-2004
C;Accession: H83778
R;Takami, H.; Nakasone, K.; Takaki, Y.; Maeno, G.; Sasaki, R.; Masui, N.; Fuji, F.; Hiran
Nucleic Acids Res. 28, 4317-4331, 2000
A;Title: Complete genome sequence of the alkaliphilic bacterium Bacillus halodurans and
A;Reference number: A83650; MUID:20512582; PMID:11058132
A;Accession: H83778
A;Status: preliminary

A;Molecule type: DNA
A;Residues: 1-14 <STO>
A;Cross-references: UNIPROT:Q9KE26; UNIPARC:UPI00000C3A89; GB:AP001510; GB:BA000004; NID
A;Experimental source: strain C-125
C;Genetics:
A;Gene: BH1032

Query Match 100.0%; Score 10; DB 2; Length 14;
Best Local Similarity 100.0%; Pred. No. 1e+03;
Matches 2; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

Qy 1 RM 2
||
Db 6 RM 7

RESULT 9
A56786
pimeloyl-CoA synthase - Bacillus sphaericus (fragment)
C;Species: Bacillus sphaericus
C;Date: 18-Aug-1995 #sequence_revision 18-Aug-1995 #text_change 18-Aug-1995
C;Accession: A56786
R;Ploux, O.; Soularue, P.; Marquet, A.; Gloeckler, R.; Lemoine, Y.
Biochem. J. 287, 685-690, 1992
A;Title: Investigation of the first step of biotin biosynthesis in Bacillus sphaericus.
A;Reference number: A56786; MUID:93075017; PMID:1445232
A;Accession: A56786
A;Status: preliminary
A;Molecule type: protein
A;Residues: 1-15 <PLO>
A;Cross-references: UNIPARC:UPI000017ACA8
A;Experimental source: Bacillus sphaericus protein overexpressed in Escherichia coli
A;Note: sequence extracted from NCBI backbone (NCBIP:117639)
C;Genetics:
A;Gene: bioW
C;Keywords: biotin biosynthesis; homodimer

Query Match 100.0%; Score 10; DB 2; Length 15;
Best Local Similarity 100.0%; Pred. No. 1.1e+03;
Matches 2; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

Qy 1 RM 2
||
Db 8 RM 9

RESULT 10
PA0020
protein QA100028 - Arabidopsis thaliana (fragment)
C;Species: Arabidopsis thaliana (mouse-ear cress)
C;Date: 30-Jun-1992 #sequence_revision 06-Jan-1995 #text_change 06-Jun-1997
C;Accession: PA0020
R;Kamo, M.; Kawakami, T.; Miyatake, N.; Tsugita, A.
submitted to JIPID, July 1994
A;Description: Separation and characterization of Arabidopsis proteins by two-dimensiona
A;Reference number: PA0001
A;Accession: PA0020
A;Molecule type: protein
A;Residues: 1-15 <KAM>
A;Cross-references: UNIPARC:UPI000017AFE7
A;Experimental source: callus

Query Match 100.0%; Score 10; DB 2; Length 15;
Best Local Similarity 100.0%; Pred. No. 1.1e+03;
Matches 2; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

Qy 1 RM 2
||
Db 7 RM 8

RESULT 11
A45096
thyrotropin-releasing hormone receptor, splice form 387 - rat (fragment)
C;Species: Rattus norvegicus (Norway rat)
C;Date: 30-Apr-1993 #sequence_revision 18-Nov-1994 #text_change 05-Nov-1999
C;Accession: A45096
R;de la Pena, P.; Delgado, L.M.; del Camino, D.; Barros, F.
J. Biol. Chem. 267, 25703-25708, 1992
A;Title: Two isoforms of the thyrotropin-releasing hormone receptor generated by alternat
A;Reference number: A45096; MUID:93100278; PMID:1334485
A;Accession: A45096
A;Status: preliminary; not compared with conceptual translation
A;Molecule type: mRNA
A;Residues: 1-15 <DE1>
A;Cross-references: UNIPARC:UPI0000170C44; GB:S51512; NID:g261982; PIDN:AAB24549.1; PID:g
A;Experimental source: GH3 anterior pituitary cells
A;Note: sequence extracted from NCBI backbone (NCBIP:120927)

Query Match 100.0%; Score 10; DB 2; Length 15;
Best Local Similarity 100.0%; Pred. No. 1.1e+03;
Matches 2; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

Qy 1 RM 2
||
Db 4 RM 5

RESULT 12
S24667
protein-tyrosine kinase (EC 2.7.1.112) kit - mouse (fragment)
C;Species: Mus musculus (house mouse)
C;Date: 03-Mar-1994 #sequence_revision 13-Mar-1997 #text_change 04-Feb-2000
C;Accession: S24667
R;Sorrentino, V.
submitted to the EMBL Data Library, May 1992
A;Reference number: S24666
A;Accession: S24667
A;Status: preliminary
A;Molecule type: DNA
A;Residues: 1-16 <SOR>
A;Cross-references: UNIPARC:UPI000016CC9E; EMBL:X65998; NID:g50433; PIDN:CAA46799.1; PID:
C;Keywords: ATP; phosphotransferase; tyrosine-specific protein kinase

Query Match 100.0%; Score 10; DB 2; Length 16;
Best Local Similarity 100.0%; Pred. No. 1.2e+03;
Matches 2; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

Qy 1 RM 2
||
Db 15 RM 16

RESULT 13
S11805
heat shock protein groEL - Comamonas acidovorans (fragment)
C;Species: Comamonas acidovorans
C;Date: 12-Feb-1993 #sequence_revision 12-Feb-1993 #text_change 09-Jul-2004
C;Accession: S11805; S22676
R;Hallett, P.; Mehlert, A.; Maxwell, A.
Mol. Microbiol. 4, 345-353, 1990
A;Title: Escherichia coli cells resistant to the DNA gyrase inhibitor, ciprofloxacin, ove
A;Reference number: S11805; MUID:90286912; PMID:1972534
A;Accession: S11805
A;Molecule type: protein
A;Residues: 1-16 <HAL>
A;Cross-references: UNIPROT:Q9R5K5; UNIPARC:UPI000017A951
A;Note: the species identification has been revised in reference S22676
R;Fowell, S.L.; Lilley, K.S.; Jones, D.; Maxwell, A.
Mol. Microbiol. 6, 1575-1576, 1992
A;Title: GroEL proteins from three Pseudomonas species.
A;Reference number: S22676; MUID:92326643; PMID:1352616
A;Accession: S22676
A;Molecule type: protein
A;Residues: 1-15 <FOW>
A;Cross-references: UNIPARC:UPI00000BED0B

A;Note: the source is designated as Pseudomonas acidovorans
C;Genetics:
A;Gene: groEL
C;Keywords: molecular chaperone

Query Match 100.0%; Score 10; DB 2; Length 16;
Best Local Similarity 100.0%; Pred. No. 1.2e+03;
Matches 2; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

Qy 1 RM 2
||
Db 14 RM 15

RESULT 14
A41170
photosystem II 6.1K protein - Chlamydomonas reinhardtii (fragment)
C;Species: Chlamydomonas reinhardtii
C;Date: 05-Jun-1992 #sequence_revision 05-Jun-1992 #text_change 09-Jul-2004
C;Accession: A41170
R;de Vitry, C.; Diner, B.A.; Popot, J.L.
J. Biol. Chem. 266, 16614-16621, 1991
A;Title: Photosystem II particles from Chlamydomonas reinhardtii. Purification, molecula
A;Reference number: A41170; MUID:91358452; PMID:1885590
A;Accession: A41170
A;Status: preliminary
A;Molecule type: protein
A;Residues: 1-16 <DEV>
A;Cross-references: UNIPROT:Q7M1J4; UNIPARC:UPI000017AF05

Query Match 100.0%; Score 10; DB 2; Length 16;
Best Local Similarity 100.0%; Pred. No. 1.2e+03;
Matches 2; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

Qy 1 RM 2
||
Db 5 RM 6

RESULT 15
S78415
ribosomal protein RL27, mitochondrial [validated] - rat (tentative sequence) (fragment)
C;Species: Rattus norvegicus (Norway rat)
C;Date: 25-Feb-1998 #sequence_revision 13-Mar-1998 #text_change 09-Jul-2004
C;Accession: S78415
R;Goldschmidt-Reisin, S.; Graack, H.R.
submitted to the Protein Sequence Database, February 1998
A;Reference number: S78411
A;Accession: S78415
A;Molecule type: protein
A;Residues: 1-16 <GOL>
A;Cross-references: UNIPROT:Q7M0C9; UNIPARC:UPI000017C9B3
A;Note: 1-Val was also found
A;Note: the protein is designated as mitochondrial ribosomal protein L27
C;Keywords: mitochondrion; protein biosynthesis; ribosome

Query Match 100.0%; Score 10; DB 2; Length 16;
Best Local Similarity 100.0%; Pred. No. 1.2e+03;
Matches 2; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

Qy 1 RM 2
||
Db 4 RM 5

Search completed: February 28, 2006, 08:54:19
Job time : 1.12197 secs

GenCore version 5.1.7

OM protein - protein search, using sw model

Run on: February 28, 2006, 08:45:31 ; Search time 0.7335 Seconds
(without alignments)
1923.730 Million cell updates/sec

Title: US-10-717-665A-44_COPY_448_449
Perfect score: 10
Sequence: 1 RM 2

Scoring table: BLOSUM62
Gapop 10.0 , Gapext 0.5

Searched: 2166443 seqs, 705528306 residues

Total number of hits satisfying chosen parameters: 2166443

Minimum DB seq length: 0
Maximum DB seq length: 2000000000

Post-processing: Minimum Match 0%
Maximum Match 100%
Listing first 45 summaries

Database : UniProt_05.80:*
1: uniprot_sprot:*
2: uniprot_trembl:*

Pred. No. is the number of results predicted by chance to have a score greater than or equal to the score of the result being printed, and is derived by analysis of the total score distribution.

SUMMARIES

Result No.	Score	% Query Match	Length	DB	ID	Description
1	10	100.0	8	2	Q5IA47_9MAGN	Q5ia47 laurus nobi
2	10	100.0	9	2	Q9UKJ6_HUMAN	Q9ukj6 homo sapien
3	10	100.0	9	2	Q4Y954_PLACH	Q4y954 plasmodium
4	10	100.0	9	2	Q5IA44_9ARAE	Q5ia44 philodendro
5	10	100.0	9	2	Q5IA45_MAHBE	Q5ia45 mahonia bea
6	10	100.0	10	2	Q7M0M6_DESDE	Q7m0m6 desulfovibr
7	10	100.0	10	2	Q9QVF1_9MURI	Q9qvf1 mus sp. pro
8	10	100.0	10	2	Q8JFE7_FICAL	Q8jfe7 ficedula al
9	10	100.0	10	2	Q8JJ33_9PASS	Q8jj33 ficedula hy
10	10	100.0	11	2	Q709B0_HUMAN	Q709b0 homo sapien
11	10	100.0	12	2	Q4VXX2_HUMAN	Q4vxx2 homo sapien
12	10	100.0	12	2	Q9AXW1_BRANA	Q9axw1 brassica na
13	10	100.0	12	2	P95606_ALCEU	P95606 alcaligenes
14	10	100.0	12	2	O50303_BACST	O50303 bacillus st
15	10	100.0	13	2	Q9TQS4_HORSE	Q9tqs4 equus cabal
16	10	100.0	13	2	Q6LCH4_RAT	Q6lch4 rattus norv
17	10	100.0	14	2	O95179_HUMAN	O95179 homo sapien
18	10	100.0	14	2	Q8TD29_HUMAN	Q8td29 homo sapien
19	10	100.0	14	2	Q93UU0_ECO57	Q93uu0 escherichia
20	10	100.0	14	2	Q9KE26_BACHD	Q9ke26 bacillus ha
21	10	100.0	14	2	Q9R1G8_RAT	Q9r1g8 rattus norv
22	10	100.0	15	2	Q16297_HUMAN	Q16297 homo sapien
23	10	100.0	15	2	Q4XKA1_PLACH	Q4xka1 plasmodium
24	10	100.0	15	2	Q9TRT3_PIG	Q9trt3 sus scrofa
25	10	100.0	15	2	Q9TRT5_PIG	Q9trt5 sus scrofa
26	10	100.0	15	2	P93515_ARATH	P93515 arabidopsis
27	10	100.0	15	2	P93516_ARATH	P93516 arabidopsis
28	10	100.0	15	2	Q5XPT1_SOLTU	Q5xpt1 solanum tub
29	10	100.0	15	2	Q7XQX8_ORYSA	Q7xqx8 oryza sativ
30	10	100.0	15	2	Q6LCK0_MOUSE	Q6lck0 mus musculu
31	10	100.0	15	2	Q35921_SALSA	Q35921 salmo salar
32	10	100.0	15	2	Q4VN45_FUGRU	Q4vn45 fugu rubrip
33	10	100.0	16	2	Q9AXV9_BRAOL	Q9axv9 brassica ol
34	10	100.0	16	2	Q9AXW0_BRACM	Q9axw0 brassica ca
35	10	100.0	16	2	Q9AXW2_BRANA	Q9axw2 brassica na
36	10	100.0	16	2	Q7M1J4_CHLRE	Q7m1j4 chlamydomon
37	10	100.0	16	2	Q7M0C9_RAT	Q7m0c9 rattus norv
38	10	100.0	16	2	Q9QW74_9MURI	Q9qw74 mus sp. hom
39	10	100.0	17	2	Q7S3Q9_NEUCR	Q7s3q9 neurospora
40	10	100.0	17	2	Q76N52_HUMAN	Q76n52 homo sapien
41	10	100.0	17	2	Q9TWY3_LEIME	Q9twy3 leishmania
42	10	100.0	17	2	Q4XMI0_PLACH	Q4xmi0 plasmodium
43	10	100.0	17	2	Q5UU76_9LAMI	Q5uu76 mentha suav
44	10	100.0	17	2	Q5UU77_9LAMI	Q5uu77 mentha spic
45	10	100.0	17	2	Q5UU78_9LAMI	Q5uu78 mentha requ

ALIGNMENTS

RESULT 1
Q5IA47_9MAGN
ID Q5IA47_9MAGN PRELIMINARY; PRT; 8 AA.
AC Q5IA47;
DT 10-MAY-2005 (TrEMBLrel. 30, Created)
DT 10-MAY-2005 (TrEMBLrel. 30, Last sequence update)
DT 10-MAY-2005 (TrEMBLrel. 30, Last annotation update)
DE NADH dehydrogenase subunit 4 (Fragment).
GN Name=nad4;
OS Laurus nobilis.
OG Mitochondrion.
OC Eukaryota; Viridiplantae; Streptophyta; Embryophyta; Tracheophyta;
OC Spermatophyta; Magnoliophyta; magnoliids; Laurales; Lauraceae; Laurus.
OX NCBI_TaxID=85223;
RN [1]
RP NUCLEOTIDE SEQUENCE.
RX PubMed=15598737; DOI=10.1073/pnas.0408336102;
RA Bergthorsson U., Richardson A.O., Young G.J., Goertzen L.R.,
RA Palmer J.D.;
RT "Massive horizontal transfer of mitochondrial genes from diverse land
RT plant donors to the basal angiosperm Amborella.";
RL Proc. Natl. Acad. Sci. U.S.A. 101:17747-17752(2004).
DR EMBL; AY832169; AAW31502.1; -; Genomic_DNA.
DR GO; GO:0005739; C:mitochondrion; IEA.
KW Mitochondrion.
FT NON_TER 1 1
FT NON_TER 8 8
SQ SEQUENCE 8 AA; 872 MW; 1B71B2C8768412C7 CRC64;

Query Match 100.0%; Score 10; DB 2; Length 8;
Best Local Similarity 100.0%; Pred. No. 2.2e+06;
Matches 2; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

```
Qy          1 RM 2
              ||
Db          3 RM 4
```

RESULT 2
Q9UKJ6_HUMAN
ID Q9UKJ6_HUMAN PRELIMINARY; PRT; 9 AA.
AC Q9UKJ6;
DT 01-MAY-2000 (TrEMBLrel. 13, Created)
DT 01-MAY-2000 (TrEMBLrel. 13, Last sequence update)
DT 01-JUN-2003 (TrEMBLrel. 24, Last annotation update)
DE Androgen receptor (Fragment).
OS Homo sapiens (Human).
OC Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;
OC Mammalia; Eutheria; Euarchontoglires; Primates; Catarrhini; Hominidae;
OC Homo.
OX NCBI_TaxID=9606;
RN [1]
RP NUCLEOTIDE SEQUENCE.
RA Zhang W., Li X.Q., Wu Q.F.;

```
RL   Submitted (JUN-1999) to the EMBL/GenBank/DDBJ databases.
DR   EMBL; AF159087; AAF04001.1; -; Genomic_DNA.
DR   GO; GO:0004872; F:receptor activity; IEA.
KW   Receptor.
FT   NON_TER       1      1
FT   NON_TER       9      9
SQ   SEQUENCE   9 AA;  1272 MW;  6F2B8415B331E684 CRC64;

  Query Match             100.0%;  Score 10;  DB 2;  Length 9;
  Best Local Similarity   100.0%;  Pred. No. 2.2e+06;
  Matches    2;  Conservative    0;  Mismatches    0;  Indels    0;  Gaps    0;

Qy          1 RM 2
               ||
Db          2 RM 3


RESULT 3
Q4Y954_PLACH
ID   Q4Y954_PLACH PRELIMINARY;      PRT;     9 AA.
AC   Q4Y954;
DT   13-SEP-2005 (TrEMBLrel. 31, Created)
DT   13-SEP-2005 (TrEMBLrel. 31, Last sequence update)
DT   13-SEP-2005 (TrEMBLrel. 31, Last annotation update)
DE   Hypothetical protein (Fragment).
GN   ORFNames=PC100056.00.0;
OS   Plasmodium chabaudi.
OC   Eukaryota; Alveolata; Apicomplexa; Haemosporida; Plasmodium.
OX   NCBI_TaxID=5825;
RN   [1]
RP   NUCLEOTIDE SEQUENCE.
RA   Hall N., Karras M., Raine J.D., Carlton J.M., Kooij T.W.A.,
RA   Berriman M., Florens L., Janssen C.S., Pain A., Christophides G.K.,
RA   James K., Rutherford K., Harris B., Harris D., Churcher C.,
RA   Quail M.A., Ormond D., Doggett J., Trueman H.E., Mendoza J.,
RA   Bidwell S.L., Rajandream M.A., Carucci D.J, Yates J.R., Kafatos F.C.,
RA   Janse C.J., Barrell B., Turner C.M.R., Waters A.P., Sinden R.S.;
RT   "A comprehensive survey of the Plasmodium life cycle by genomic,
RT   transcriptomic, and proteomic analyses.";
RL   Science 307:82-86(2005).
CC   -!- CAUTION: The sequence shown here is derived from an
CC       EMBL/GenBank/DDBJ whole genome shotgun (WGS) entry which is
CC       preliminary data.
DR   EMBL; CAAJ01000035; CAH74304.1; -; Genomic_DNA.
KW   Hypothetical protein.
FT   NON_TER       1      1
SQ   SEQUENCE   9 AA;  1124 MW;  0528840452C699C0 CRC64;

  Query Match             100.0%;  Score 10;  DB 2;  Length 9;
  Best Local Similarity   100.0%;  Pred. No. 2.2e+06;
  Matches    2;  Conservative    0;  Mismatches    0;  Indels    0;  Gaps    0;

Qy          1 RM 2
               ||
Db          7 RM 8


RESULT 4
Q5IA44_9ARAE
ID   Q5IA44_9ARAE PRELIMINARY;      PRT;     9 AA.
AC   Q5IA44;
DT   10-MAY-2005 (TrEMBLrel. 30, Created)
DT   10-MAY-2005 (TrEMBLrel. 30, Last sequence update)
DT   10-MAY-2005 (TrEMBLrel. 30, Last annotation update)
DE   NADH dehydrogenase subunit 4 (Fragment).
GN   Name=nad4;
OS   Philodendron oxycardium.
OG   Mitochondrion.
OC   Eukaryota; Viridiplantae; Streptophyta; Embryophyta; Tracheophyta;
OC   Spermatophyta; Magnoliophyta; Liliopsida; Araceae; Aroideae;
OC   Philodendreae; Philodendron.
OX   NCBI_TaxID=71614;
RN   [1]
RP   NUCLEOTIDE SEQUENCE.
RX   PubMed=15598737; DOI=10.1073/pnas.0408336102;
RA   Bergthorsson U., Richardson A.O., Young G.J., Goertzen L.R.,
RA   Palmer J.D.;
RT   "Massive horizontal transfer of mitochondrial genes from diverse land
RT   plant donors to the basal angiosperm Amborella.";
RL   Proc. Natl. Acad. Sci. U.S.A. 101:17747-17752(2004).
DR   EMBL; AY832172; AAW31505.1; -; Genomic_DNA.
DR   GO; GO:0005739; C:mitochondrion; IEA.
KW   Mitochondrion.
FT   NON_TER       1      1
FT   NON_TER       9      9
SQ   SEQUENCE   9 AA;  996 MW;  D35B7772C8768412 CRC64;

  Query Match             100.0%;  Score 10;  DB 2;  Length 9;
  Best Local Similarity   100.0%;  Pred. No. 2.2e+06;
  Matches    2;  Conservative    0;  Mismatches    0;  Indels    0;  Gaps    0;

Qy          1 RM 2
               ||
Db          3 RM 4


RESULT 5
Q5IA45_MAHBE
ID   Q5IA45_MAHBE PRELIMINARY;      PRT;     9 AA.
AC   Q5IA45;
DT   10-MAY-2005 (TrEMBLrel. 30, Created)
DT   10-MAY-2005 (TrEMBLrel. 30, Last sequence update)
DT   10-MAY-2005 (TrEMBLrel. 30, Last annotation update)
DE   NADH dehydrogenase subunit 4 (Fragment).
GN   Name=nad4;
OS   Mahonia bealei (Leatherleaf mahonia).
OG   Mitochondrion.
OC   Eukaryota; Viridiplantae; Streptophyta; Embryophyta; Tracheophyta;
OC   Spermatophyta; Magnoliophyta; eudicotyledons; Ranunculales;
OC   Berberidaceae; Mahonia.
OX   NCBI_TaxID=13601;
RN   [1]
RP   NUCLEOTIDE SEQUENCE.
RX   PubMed=15598737; DOI=10.1073/pnas.0408336102;
RA   Bergthorsson U., Richardson A.O., Young G.J., Goertzen L.R.,
RA   Palmer J.D.;
RT   "Massive horizontal transfer of mitochondrial genes from diverse land
RT   plant donors to the basal angiosperm Amborella.";
RL   Proc. Natl. Acad. Sci. U.S.A. 101:17747-17752(2004).
DR   EMBL; AY832171; AAW31504.1; -; Genomic_DNA.
DR   GO; GO:0005739; C:mitochondrion; IEA.
KW   Mitochondrion.
FT   NON_TER       1      1
FT   NON_TER       9      9
SQ   SEQUENCE   9 AA;  1000 MW;  D35B71B2C8768412 CRC64;

  Query Match             100.0%;  Score 10;  DB 2;  Length 9;
  Best Local Similarity   100.0%;  Pred. No. 2.2e+06;
  Matches    2;  Conservative    0;  Mismatches    0;  Indels    0;  Gaps    0;

Qy          1 RM 2
               ||
Db          3 RM 4


RESULT 6
Q7M0M6_DESDE
ID   Q7M0M6_DESDE PRELIMINARY;      PRT;    10 AA.
AC   Q7M0M6;
DT   01-MAR-2004 (TrEMBLrel. 26, Created)
DT   01-MAR-2004 (TrEMBLrel. 26, Last sequence update)
DT   01-MAR-2004 (TrEMBLrel. 26, Last annotation update)
DE   Cytochrome c553 (Fragment).
OS   Desulfovibrio desulfuricans.
OC   Bacteria; Proteobacteria; Deltaproteobacteria; Desulfovibrionales;
OC   Desulfovibrionaceae; Desulfovibrio.
```

```
OX   NCBI_TaxID=876;
RN   [1]
RP   PROTEIN SEQUENCE.
RX   MEDLINE=98102811; PubMed=9439638; DOI=10.1006/bbrc.1997.7852;
RA   Aubert C., Leroy G., Bianco P., Forest E., Bruschi M., Dolla A.;
RT   "Characterization of the cytochromes C from Desulfovibrio
RT   desulfuricans G201.";
RL   Biochem. Biophys. Res. Commun. 242:213-218(1998).
DR   PIR; PC4442; PC4442.
FT   NON_TER       1      1
FT   NON_TER      10     10
SQ   SEQUENCE   10 AA;  1112 MW;  87C51E5735BDD87A CRC64;

  Query Match             100.0%;  Score 10;  DB 2;  Length 10;
  Best Local Similarity   100.0%;  Pred. No. 3.8e+03;
  Matches    2;  Conservative    0;  Mismatches    0;  Indels    0;  Gaps    0;

Qy          1 RM 2
              ||
Db          9 RM 10


RESULT 7
Q9QVF1_9MURI
ID   Q9QVF1_9MURI PRELIMINARY;      PRT;    10 AA.
AC   Q9QVF1;
DT   01-MAY-2000 (TrEMBLrel. 13, Created)
DT   01-MAY-2000 (TrEMBLrel. 13, Last sequence update)
DT   01-JUN-2002 (TrEMBLrel. 21, Last annotation update)
DE   Protamine MP2 (Fragment).
OS   Mus sp.
OC   Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;
OC   Mammalia; Eutheria; Euarchontoglires; Glires; Rodentia; Sciurognathi;
OC   Muridae; Murinae; Mus.
OX   NCBI_TaxID=10095;
RN   [1]
RP   PROTEIN SEQUENCE.
RX   MEDLINE=92174934; PubMed=1541289;
RA   Chauviere M., Martinage A., Debarle M., Sautiere P., Chevaillier P.;
RT   "Molecular characterization of six intermediate proteins in the
RT   processing of mouse protamine P2 precursor.";
RL   Eur. J. Biochem. 204:759-765(1992).
FT   NON_TER       1      1
FT   NON_TER      10     10
SQ   SEQUENCE   10 AA;  1280 MW;  01DD2975A406841B CRC64;

  Query Match             100.0%;  Score 10;  DB 2;  Length 10;
  Best Local Similarity   100.0%;  Pred. No. 3.8e+03;
  Matches    2;  Conservative    0;  Mismatches    0;  Indels    0;  Gaps    0;

Qy          1 RM 2
              ||
Db          4 RM 5


RESULT 8
Q8JFE7_FICAL
ID   Q8JFE7_FICAL PRELIMINARY;      PRT;    10 AA.
AC   Q8JFE7;
DT   01-OCT-2002 (TrEMBLrel. 22, Created)
DT   01-OCT-2002 (TrEMBLrel. 22, Last sequence update)
DT   01-FEB-2005 (TrEMBLrel. 29, Last annotation update)
DE   Myelin proteolipid protein (Fragment).
OS   Ficedula albicollis (Collared flycatcher).
OC   Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;
OC   Archosauria; Aves; Neognathae; Passeriformes; Muscicapidae; Ficedula.
OX   NCBI_TaxID=59894;
RN   [1]
RP   NUCLEOTIDE SEQUENCE.
RX   MEDLINE=21918460; PubMed=11918793;
RA   Primmer C.R., Borge T., Lindell J., Saetre G.P.;
RT   "Single-nucleotide polymorphism characterization in species with
RT   limited available sequence information: high nucleotide diversity
RT   revealed in the avian genome.";
RL   Mol. Ecol. 11:603-612(2002).
DR   EMBL; AF454217; AAM22903.1; -; Genomic_DNA.
DR   EMBL; AF454218; AAM22904.1; -; Genomic_DNA.
FT   NON_TER       1      1
FT   NON_TER      10     10
SQ   SEQUENCE   10 AA;  1206 MW;  1FAAC9676732C86B CRC64;

  Query Match             100.0%;  Score 10;  DB 2;  Length 10;
  Best Local Similarity   100.0%;  Pred. No. 3.8e+03;
  Matches    2;  Conservative    0;  Mismatches    0;  Indels    0;  Gaps    0;

Qy          1 RM 2
              ||
Db          1 RM 2


RESULT 9
Q8JJ33_9PASS
ID   Q8JJ33_9PASS PRELIMINARY;      PRT;    10 AA.
AC   Q8JJ33;
DT   01-OCT-2002 (TrEMBLrel. 22, Created)
DT   01-OCT-2002 (TrEMBLrel. 22, Last sequence update)
DT   01-OCT-2002 (TrEMBLrel. 22, Last annotation update)
DE   Myelin proteolipid protein (Fragment).
OS   Ficedula hypoleuca.
OC   Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;
OC   Archosauria; Aves; Neognathae; Passeriformes; Muscicapidae; Ficedula.
OX   NCBI_TaxID=46689;
RN   [1]
RP   NUCLEOTIDE SEQUENCE.
RX   MEDLINE=21918460; PubMed=11918793;
RA   Primmer C.R., Borge T., Lindell J., Saetre G.P.;
RT   "Single-nucleotide polymorphism characterization in species with
RT   limited available sequence information: high nucleotide diversity
RT   revealed in the avian genome.";
RL   Mol. Ecol. 11:603-612(2002).
DR   EMBL; AF454216; AAM22902.1; -; Genomic_DNA.
FT   NON_TER       1      1
FT   NON_TER      10     10
SQ   SEQUENCE   10 AA;  1206 MW;  1FAAC9676732C86B CRC64;

  Query Match             100.0%;  Score 10;  DB 2;  Length 10;
  Best Local Similarity   100.0%;  Pred. No. 3.8e+03;
  Matches    2;  Conservative    0;  Mismatches    0;  Indels    0;  Gaps    0;

Qy          1 RM 2
              ||
Db          1 RM 2


RESULT 10
Q709B0_HUMAN
ID   Q709B0_HUMAN PRELIMINARY;      PRT;    11 AA.
AC   Q709B0;
DT   05-JUL-2004 (TrEMBLrel. 27, Created)
DT   05-JUL-2004 (TrEMBLrel. 27, Last sequence update)
DT   01-FEB-2005 (TrEMBLrel. 29, Last annotation update)
DE   Transcription factor EB (Fragment).
GN   Name=TFEB;
OS   Homo sapiens (Human).
OC   Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;
OC   Mammalia; Eutheria; Euarchontoglires; Primates; Catarrhini; Hominidae;
OC   Homo.
OX   NCBI_TaxID=9606;
RN   [1]
RP   NUCLEOTIDE SEQUENCE.
RC   TISSUE=Brain;
RA   Kuiper R.P., Schepens M., Thijssen J., Schoenmakers E.F.P.M.,
RA   Geurts van Kessel A.;
RT   "Regulation of the MITF/TFE bHLH-LZ transcription factors through
RT   restricted spatial expression and alternative splicing of functional
RT   domains.";
```

```
RL   Nucleic Acids Res. 32:2315-2322(2004).
DR   EMBL; AJ608789; CAE77674.1; -; mRNA.
DR   EMBL; AJ608792; CAE77677.1; -; mRNA.
DR   EMBL; AJ608788; CAE77673.1; -; mRNA.
FT   NON_TER      11     11
SQ   SEQUENCE   11 AA;  1276 MW;  7BB98CBD07287044 CRC64;

  Query Match             100.0%;  Score 10;  DB 2;  Length 11;
  Best Local Similarity   100.0%;  Pred. No. 4.2e+03;
  Matches    2;  Conservative    0;  Mismatches    0;  Indels    0;  Gaps    0;

Qy          1 RM 2
              ||
Db          8 RM 9


RESULT 11
Q4VXX2_HUMAN
ID   Q4VXX2_HUMAN PRELIMINARY;      PRT;    12 AA.
AC   Q4VXX2;
DT   13-SEP-2005 (TrEMBLrel. 31, Created)
DT   13-SEP-2005 (TrEMBLrel. 31, Last sequence update)
DT   13-SEP-2005 (TrEMBLrel. 31, Last annotation update)
DE   Transcription factor EB (Fragment).
GN   Name=TFEB; ORFNames=RP4-696P19.3-012;
OS   Homo sapiens (Human).
OC   Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;
OC   Mammalia; Eutheria; Euarchontoglires; Primates; Catarrhini; Hominidae;
OC   Homo.
OX   NCBI_TaxID=9606;
RN   [1]
RP   NUCLEOTIDE SEQUENCE.
RA   Tracey A.;
RL   Submitted (MAY-2005) to the EMBL/GenBank/DDBJ databases.
DR   EMBL; AL035588; CAI95733.1; -; Genomic_DNA.
FT   NON_TER      12     12
SQ   SEQUENCE   12 AA;  1407 MW;  0C4BB98CBD072870 CRC64;

  Query Match             100.0%;  Score 10;  DB 2;  Length 12;
  Best Local Similarity   100.0%;  Pred. No. 4.6e+03;
  Matches    2;  Conservative    0;  Mismatches    0;  Indels    0;  Gaps    0;

Qy          1 RM 2
              ||
Db          8 RM 9


RESULT 12
Q9AXW1_BRANA
ID   Q9AXW1_BRANA PRELIMINARY;      PRT;    12 AA.
AC   Q9AXW1;
DT   01-JUN-2001 (TrEMBLrel. 17, Created)
DT   01-JUN-2001 (TrEMBLrel. 17, Last sequence update)
DT   01-JUN-2001 (TrEMBLrel. 17, Last annotation update)
DE   Phytochrome A (Fragment).
GN   Name=BN-PHYA-2;
OS   Brassica napus (Rape).
OC   Eukaryota; Viridiplantae; Streptophyta; Embryophyta; Tracheophyta;
OC   Spermatophyta; Magnoliophyta; eudicotyledons; core eudicots; rosids;
OC   eurosids II; Brassicales; Brassicaceae; Brassica.
OX   NCBI_TaxID=3708;
RN   [1]
RP   NUCLEOTIDE SEQUENCE.
RX   PubMed=12582899;
RA   Fourmann M., Barret P., Froger N., Baron C., Charlot F., Delourme R.,
RA   Brunel D.;
RT   "From Arabidopsis thaliana to Brassica napus: development of amplified
RT   consensus genetic markers (ACGM) for construction of a gene map.";
RL   Theor. Appl. Genet. 105:1196-1206(2002).
DR   EMBL; AF229414; AAK00686.1; -; Genomic_DNA.
DR   GO; GO:0008020; F:G-protein coupled photoreceptor activity; IEA.
DR   GO; GO:0009585; P:red, far-red light phototransduction; IEA.
KW   Phytochrome.
FT   NON_TER       1      1
FT   NON_TER      12     12
SQ   SEQUENCE   12 AA;  1442 MW;  7500ECB95A172684 CRC64;

  Query Match             100.0%;  Score 10;  DB 2;  Length 12;
  Best Local Similarity   100.0%;  Pred. No. 4.6e+03;
  Matches    2;  Conservative    0;  Mismatches    0;  Indels    0;  Gaps    0;

Qy          1 RM 2
              ||
Db          5 RM 6


RESULT 13
P95606_ALCEU
ID   P95606_ALCEU PRELIMINARY;      PRT;    12 AA.
AC   P95606;
DT   01-MAY-1997 (TrEMBLrel. 03, Created)
DT   01-MAY-1997 (TrEMBLrel. 03, Last sequence update)
DT   01-MAR-2001 (TrEMBLrel. 16, Last annotation update)
DE   Nickel permease (Fragment).
GN   Name=hoxN;
OS   Alcaligenes eutrophus (Ralstonia eutropha).
OC   Bacteria; Proteobacteria; Betaproteobacteria; Burkholderiales;
OC   Burkholderiaceae; Cupriavidus.
OX   NCBI_TaxID=106590;
RN   [1]
RP   NUCLEOTIDE SEQUENCE.
RC   STRAIN=H16;
RX   MEDLINE=91131629; PubMed=1847142;
RA   Eitinger T., Friedrich B.;
RT   "Cloning, nucleotide sequence, and heterologous expression of a high-
RT   affinity nickel transport gene from Alcaligenes eutrophus.";
RL   J. Biol. Chem. 266:3222-3227(1991).
RN   [2]
RP   NUCLEOTIDE SEQUENCE.
RC   STRAIN=H16;
RA   Lenz O.;
RL   Submitted (DEC-1996) to the EMBL/GenBank/DDBJ databases.
DR   EMBL; U82564; AAB49367.1; -; Genomic_DNA.
FT   NON_TER      12     12
SQ   SEQUENCE   12 AA;  1367 MW;  478C45052BC87DD7 CRC64;

  Query Match             100.0%;  Score 10;  DB 2;  Length 12;
  Best Local Similarity   100.0%;  Pred. No. 4.6e+03;
  Matches    2;  Conservative    0;  Mismatches    0;  Indels    0;  Gaps    0;

Qy          1 RM 2
              ||
Db          9 RM 10


RESULT 14
O50303_BACST
ID   O50303_BACST PRELIMINARY;      PRT;    12 AA.
AC   O50303;
DT   01-JUN-1998 (TrEMBLrel. 06, Created)
DT   01-JUN-1998 (TrEMBLrel. 06, Last sequence update)
DT   01-OCT-2003 (TrEMBLrel. 25, Last annotation update)
DE   Hypothetical protein ORF2 (Fragment).
GN   Name=ORF2;
OS   Bacillus stearothermophilus.
OC   Bacteria; Firmicutes; Bacillales; Bacillaceae; Geobacillus.
OX   NCBI_TaxID=1422;
RN   [1]
RP   NUCLEOTIDE SEQUENCE.
RC   STRAIN=CCM 2184;
RX   MEDLINE=20194845; PubMed=10732707;
RA   Vlaskova H., Krasny L., Fucik V., Jonak J.;
RT   "The pyrAb gene coding for the large subunit of carbamoylphosphate
RT   synthetase from Bacillus stearothermophilus: molecular cloning and
RT   functional characterization.";
RL   Folia Biol. (Praha) 44:163-172(1998).
```

```
DR   EMBL; AJ001805; CAA05021.1; -; Genomic_DNA.
DR   PIR; T44420; T44420.
KW   Hypothetical protein.
FT   NON_TER      12     12
SQ   SEQUENCE   12 AA;  1379 MW;  70087CB0E8A6840B CRC64;

  Query Match             100.0%;  Score 10;  DB 2;  Length 12;
  Best Local Similarity   100.0%;  Pred. No. 4.6e+03;
  Matches    2;  Conservative    0;  Mismatches    0;  Indels    0;  Gaps    0;

Qy          1 RM 2
              ||
Db          6 RM 7


RESULT 15
Q9TQS4_HORSE
ID   Q9TQS4_HORSE PRELIMINARY;      PRT;    13 AA.
AC   Q9TQS4;
DT   01-MAY-2000 (TrEMBLrel. 13, Created)
DT   01-MAY-2000 (TrEMBLrel. 13, Last sequence update)
DT   01-FEB-2005 (TrEMBLrel. 29, Last annotation update)
DE   Transferrin (Fragment).
OS   Equus caballus (Horse).
OC   Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;
OC   Mammalia; Eutheria; Laurasiatheria; Perissodactyla; Equidae; Equus.
OX   NCBI_TaxID=9796;
RN   [1]
RP   NUCLEOTIDE SEQUENCE.
RA   Giffard J.M., Brandon R.B., Bell T.K.;
RL   Submitted (SEP-1999) to the EMBL/GenBank/DDBJ databases.
DR   EMBL; AF185755; AAF05469.1; -; Genomic_DNA.
DR   EMBL; AF185754; AAF05468.1; -; Genomic_DNA.
FT   NON_TER       1      1
FT   NON_TER      13     13
SQ   SEQUENCE   13 AA;  1528 MW;  D379D35F2EA2840D CRC64;

  Query Match             100.0%;  Score 10;  DB 2;  Length 13;
  Best Local Similarity   100.0%;  Pred. No. 5e+03;
  Matches    2;  Conservative    0;  Mismatches    0;  Indels    0;  Gaps    0;

Qy          1 RM 2
              ||
Db          7 RM 8


Search completed: February 28, 2006, 08:53:04
Job time : 2.7335 secs
```

This Page Blank (uspto)

GenCore version 5.1.7

OM protein - protein search, using sw model

Run on: February 28, 2006, 08:37:25 ; Search time 36.9925 Seconds
(without alignments)
1282.772 Million cell updates/sec

Title: US-10-717-665A-44_COPY_532_639
Perfect score: 551
Sequence: 1 KLAFRACRLKKKAQYEANKV.....TAEGNPTGGLVGLRIPTSKV 108

Scoring table: BLOSUM62
Gapop 10.0 , Gapext 0.5

Searched: 2443163 seqs, 439378781 residues

Total number of hits satisfying chosen parameters: 2443163

Minimum DB seq length: 0
Maximum DB seq length: 2000000000

Post-processing: Minimum Match 0%
Maximum Match 100%
Listing first 45 summaries

Database : A_Geneseq_21:*
1: geneseqp1980s:*
2: geneseqp1990s:*
3: geneseqp2000s:*
4: geneseqp2001s:*
5: geneseqp2002s:*
6: geneseqp2003as:*
7: geneseqp2003bs:*
8: geneseqp2004s:*
9: geneseqp2005s:*

Pred. No. is the number of results predicted by chance to have a score greater than or equal to the score of the result being printed, and is derived by analysis of the total score distribution.

SUMMARIES

Result No.	Score	% Query Match	Length	DB	ID	Description
1	551	100.0	639	7	ADK65805	Adk65805 Angiogene
2	543	98.5	296	8	ADO20373	Ado20373 Human PRO
3	202	36.7	814	4	ABB68374	Abb68374 Drosophil
4	200.5	36.4	160	4	ABG29275	Abg29275 Novel hum
5	82	14.9	1124	6	ABU11886	Abu11886 Human ABC
6	78.5	14.2	281	6	ABU19277	Abu19277 Protein e
7	75.5	13.7	345	6	ABU01541	Abu01541 S. pneumo
8	75.5	13.7	642	8	ADM92163	Adm92163 S pneumon
9	75.5	13.7	1175	8	ADS44237	Ads44237 Bacterial
10	74.5	13.5	1124	8	ADS30935	Ads30935 Bacterial
11	73.5	13.3	625	6	ABU43663	Abu43663 Protein e
12	73.5	13.3	1104	8	ADS30054	Ads30054 Bacterial
13	73.5	13.3	1495	2	AAW18226	Aaw18226 Transcrip
14	73.5	13.3	2000	8	ADN04305	Adn04305 Antipsori
15	73.5	13.3	2507	6	ABU61812	Abu61812 Human nuc
16	73.5	13.3	2507	8	ADG86291	Adg86291 Human SMR
17	73.5	13.3	2507	8	ADQ18921	Adq18921 Human sof
18	73.5	13.3	2514	8	ADS88212	Ads88212 Human pro
19	73.5	13.3	2517	7	ABM85413	Abm85413 Human pro
20	73.5	13.3	2517	8	ADG86299	Adg86299 Human SMR
21	73.5	13.3	2518	3	AAB40574	Aab40574 Human ORF
22	72.5	13.2	510	4	AAB48935	Aab48935 Brassica
23	72.5	13.2	642	8	ADK47294	Adk47294 Streptoco
24	72.5	13.2	651	8	ADR94210	Adr94210 Novel S.
25	72.5	13.2	651	9	AEA58080	Aea58080 Streptoco
26	72	13.1	861	8	ADN20856	Adn20856 Bacterial
27	71	12.9	505	4	ABB65505	Abb65505 Drosophil
28	71	12.9	825	8	ADN26279	Adn26279 Bacterial
29	71	12.9	1355	8	ADN22598	Adn22598 Bacterial
30	71	12.9	1355	8	ADN22599	Adn22599 Bacterial
31	70.5	12.8	667	7	ADK64278	Adk64278 Disease.t
32	70	12.7	487	4	AAG67384	Aag67384 Amino aci
33	70	12.7	597	8	ADN20612	Adn20612 Bacterial
34	70	12.7	640	8	ADV16175	Adv16175 X. pelarg
35	70	12.7	846	4	AAG67386	Aag67386 Amino aci
36	69.5	12.6	1105	8	ADN20163	Adn20163 Bacterial
37	69	12.5	152	9	ADY86677	Ady86677 Human GRP
38	69	12.5	188	9	ADY86676	Ady86676 Human GRP
39	69	12.5	190	8	ADN72115	Adn72115 Rice prot
40	69	12.5	194	8	ADN99816	Adn99816 Novel hum
41	69	12.5	220	5	ADF77333	Adf77333 Heat shoc
42	69	12.5	223	5	AAE23607	Aae23607 Human end
43	69	12.5	282	5	ABP26389	Abp26389 Streptoco
44	69	12.5	282	6	ABU46647	Abu46647 Protein e
45	69	12.5	287	8	ADN99815	Adn99815 Novel hum

ALIGNMENTS

RESULT 1
ADK65805
ID ADK65805 standard; protein; 639 AA.
XX
AC ADK65805;
XX
DT 06-MAY-2004 (first entry)
XX
DE Angiogenesis-differentially expressed protein ANH0757.
XX
KW cytostatic; cardiant; vasotropic; antiarteriosclerotic;
KW angiogenesis inhibitor; angiogenesis stimulator; angiogenic index;
KW gene expression; cancer; coronary artery disease; myocardial ischemia;
KW coronary arteriosclerosis; forensic medicine.
XX
OS Homo sapiens.
XX
PN WO2003066831-A2.
XX
PD 14-AUG-2003.
XX
PF 07-FEB-2003; 2003WO-US003848.
XX
PR 07-FEB-2002; 2002US-00067482.
PR 10-JUN-2002; 2002US-00164595.
PR 16-AUG-2002; 2002US-0403649P.
PR 03-JAN-2003; 2003US-0437746P.
XX
PA (ORIG-) ORIGENE TECHNOLOGIES INC.
XX
PI Sun Z, Li X, Kovacs KF, Fan W, Jay G;
XX
DR WPI; 2003-731502/69.
DR N-PSDB; ADK65804.
XX
PT Determining the angiogenic index of a tissue or cell sample using
PT expression levels of differentially expressed genes, useful for
PT diagnosing or treating cancer, coronary artery disease, myocardial
PT ischemia and/or arteriosclerosis.
XX
PS Claim 23; SEQ ID NO 44; 296pp; English.
XX
CC The invention relates to a method of determining the angiogenic index of
CC a tissue or cell sample comprising assessing, in a sample, the expression
CC levels of one or more differentially-expressed gene from any of 34 DNA
CC sequences, given in the specification, where the levels are indicative of
CC the angiogenic index. The methods and compositions of the present

CC invention are useful for diagnosing, preventing and/or treating cancer,
CC coronary artery disease, myocardial ischemia or coronary
CC arteriosclerosis. They can also be used in research, drug discovery and
CC forensic medicine involving angiogenesis. This sequence corresponds to
CC one of the differentially expressed proteins of the invention.
XX
SQ Sequence 639 AA;

Query Match 100.0%; Score 551; DB 7; Length 639;
Best Local Similarity 100.0%; Pred. No. 1.1e-53;
Matches 108; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

```
Qy          1 KLAFRACRLKKKAQYEANKVKLWGLNTEYDNLLFVINSIKQEIVNRVQNPRDERGPNMGQ 60
              ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db        532 KLAFRACRLKKKAQYEANKVKLWGLNTEYDNLLFVINSIKQEIVNRVQNPRDERGPNMGQ 591

Qy         61 KLEILIKDTLGLPVAGQTSEFVNQVLEKTAEGNPTGGLVGLRIPTSKV 108
              ||||||||||||||||||||||||||||||||||||||||||||||||||
Db        592 KLEILIKDTLGLPVAGQTSEFVNQVLEKTAEGNPTGGLVGLRIPTSKV 639
```

RESULT 2
ADO20373
ID ADO20373 standard; protein; 296 AA.
XX
AC ADO20373;
XX
DT 12-AUG-2004 (first entry)
XX
DE Human PRO polypeptide #633.
XX
KW Human; PRO; immune related disorder; systemic lupus erythematosus;
KW rheumatoid arthritis; osteoarthritis; juvenile chronic arthritis;
KW systemic sclerosis; Sjogren's syndrome; vasculitis; sarcoidosis;
KW autoimmune haemolytic anaemia; autoimmune thrombocytopenia; thyroiditis;
KW diabetes mellitus; renal disease; demyelinating disease;
KW central nervous system; peripheral nervous system;
KW demyelinating polyneuropathy; Guillain-Barre syndrome;
KW chronic inflammatory demyelinating polyneuropathy.
XX
OS Homo sapiens.
XX
PN WO2004043361-A2.
XX
PD 27-MAY-2004.
XX
PF 06-NOV-2003; 2003WO-US035268.
XX
PR 08-NOV-2002; 2002US-0425235P.
XX
PA (GETH) GENENTECH INC.
XX
PI Fong S, Dennis K, Clark H, Chiu H, Schoenfeld J, Williams PM;
PI Wood WI, Wu TD;
XX
DR WPI; 2004-420067/39.
DR N-PSDB; ADO20372.
XX
PT Novel PRO polypeptide e.g., PRO69614, PRO71106, or PRO86388 useful for
PT treating an immune related disorder such as systemic lupus erythematosus,
PT rheumatoid arthritis, osteoarthritis, juvenile chronic arthritis or
PT spondyloarthropathy.
XX
PS Claim 7; SEQ ID NO 1266; 1731pp; English.
XX
CC The invention relates to human PRO polypeptides and the polynucleotides
CC encoding them. The polypeptides and polynucleotides are useful for
CC treating and diagnosing immune related disorders in mammals. The immune
CC related disorders include systemic lupus erythematosus, rheumatoid
CC arthritis, osteoarthritis, juvenile chronic arthritis, systemic
CC sclerosis, Sjogren's syndrome, vasculitis, sarcoidosis, autoimmune
CC haemolytic anaemia, autoimmune thrombocytopenia, thyroiditis, diabetes
CC mellitus, immune-mediated renal disease, demyelinating diseases of the
CC central or peripheral nervous system, demyelinating polyneuropathy,
CC Guillain-Barre syndrome and chronic inflammatory demyelinating
CC polyneuropathy. This sequence represents a human PRO polypeptide of the
CC invention.
XX
SQ Sequence 296 AA;

Query Match 98.5%; Score 543; DB 8; Length 296;
Best Local Similarity 99.1%; Pred. No. 3.5e-53;
Matches 107; Conservative 0; Mismatches 1; Indels 0; Gaps 0;

```
Qy          1 KLAFRACRLKKKAQYEANKVKLWGLNTEYDNLLFVINSIKQEIVNRVQNPRDERGPNMGQ 60
              ||| ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db        189 KLASRACRLKKKAQYEANKVKLWGLNTEYDNLLFVINSIKQEIVNRVQNPRDERGPNMGQ 248

Qy         61 KLEILIKDTLGLPVAGQTSEFVNQVLEKTAEGNPTGGLVGLRIPTSKV 108
              ||||||||||||||||||||||||||||||||||||||||||||||||||
Db        249 KLEILIKDTLGLPVAGQTSEFVNQVLEKTAEGNPTGGLVGLRIPTSKV 296
```

RESULT 3
ABB68374
ID ABB68374 standard; protein; 814 AA.
XX
AC ABB68374;
XX
DT 26-MAR-2002 (first entry)
XX
DE Drosophila melanogaster polypeptide SEQ ID NO 31914.
XX
KW Drosophila; developmental biology; cell signalling; insecticide;
KW pharmaceutical.
XX
OS Drosophila melanogaster.
XX
PN WO200171042-A2.
XX
PD 27-SEP-2001.
XX
PF 23-MAR-2001; 2001WO-US009231.
XX
PR 23-MAR-2000; 2000US-0191637P.
PR 11-JUL-2000; 2000US-00614150.
XX
PA (PEKE) PE CORP NY.
XX
PI Venter JC, Adams M, Li PWD, Myers EW;
XX
DR WPI; 2001-656860/75.
DR N-PSDB; ABL12477.
XX
PT New isolated nucleic acid detection reagent for detecting 1000 or more
PT genes from Drosophila and for elucidating cell signaling and cell-cell
PT interactions.
XX
PS Disclosure; SEQ ID NO 31914; 21pp + Sequence Listing; English.
XX
CC The invention relates to an isolated nucleic acid detection reagent
CC capable of detecting 1000 or more genes from Drosophila. The invention is
CC useful in developmental biology and in elucidating cell signalling and
CC cell-cell interactions in higher eukaryotes for the development of
CC insecticides, therapeutics and pharmaceutical drugs. The invention
CC discloses genomic DNA sequences (ABL16176-ABL30511), expressed DNA
CC sequences (ABL01840-ABL16175) and the encoded proteins (ABB57737-
CC ABB72072). The sequence data for this patent did not form part of the
CC printed specification, but was obtained in electronic format directly
CC from WIPO at ftp.wipo.int/pub/published_pct_sequences
XX
SQ Sequence 814 AA;

Query Match 36.7%; Score 202; DB 4; Length 814;
Best Local Similarity 43.6%; Pred. No. 1e-13;
Matches 48; Conservative 17; Mismatches 31; Indels 14; Gaps 2;

```
Qy          1 KLAFRACRLKKKAQYEANKVKLWGLNTEYDNLLFVINSIKQEIVNRVQNPRDERGPNMGQ 60
              ||| |||||||||:||||:||:||  |:  |:  |  :|| :| :       |  |:|:
Db        708 KLASRACRLKKKAQHEANKIKLPGLEIEHKRLMNGIAELKQALVVK------HRTKNLGE 761

Qy         61 KLE--------ILIKDTLGLPVAGQTSEFVNQVLEKTAEGNPTGGLVGLR 102
                |        |    : |: :|| :::|||:|||    | || |||  ||
Db        762 STEEVDQQIARIYATASSGIRIAGGSTDFVNKVLENMRGGMPNGGLEELR 811


RESULT 4
ABG29275
ID   ABG29275 standard; protein; 160 AA.
XX
AC   ABG29275;
XX
DT   18-FEB-2002  (first entry)
XX
DE   Novel human diagnostic protein #29266.
XX
KW   Human; chromosome mapping; gene mapping; gene therapy; forensic;
KW   food supplement; medical imaging; diagnostic; genetic disorder.
XX
OS   Homo sapiens.
XX
PN   WO200175067-A2.
XX
PD   11-OCT-2001.
XX
PF   30-MAR-2001; 2001WO-US008631.
XX
PR   31-MAR-2000; 2000US-00540217.
PR   23-AUG-2000; 2000US-00649167.
XX
PA   (HYSE-) HYSEQ INC.
XX
PI   Drmanac RT,  Liu C,  Tang YT;
XX
DR   WPI; 2001-639362/73.
DR   N-PSDB; AAS93462.
XX
PT   New isolated polynucleotide and encoded polypeptides, useful in
PT   diagnostics, forensics, gene mapping, identification of mutations
PT   responsible for genetic disorders or other traits and to assess
PT   biodiversity.
XX
PS   Claim 20; SEQ ID NO 59634; 103pp; English.
XX
CC   The invention relates to isolated polynucleotide (I) and polypeptide (II)
CC   sequences. (I) is useful as hybridisation probes, polymerase chain
CC   reaction (PCR) primers, oligomers, and for chromosome and gene mapping,
CC   and in recombinant production of (II). The polynucleotides are also used
CC   in diagnostics as expressed sequence tags for identifying expressed
CC   genes. (I) is useful in gene therapy techniques to restore normal
CC   activity of (II) or to treat disease states involving (II). (II) is
CC   useful for generating antibodies against it, detecting or quantitating a
CC   polypeptide in tissue, as molecular weight markers and as a food
CC   supplement. (II) and its binding partners are useful in medical imaging
CC   of sites expressing (II). (I) and (II) are useful for treating disorders
CC   involving aberrant protein expression or biological activity. The
CC   polypeptide and polynucleotide sequences have applications in
CC   diagnostics, forensics, gene mapping, identification of mutations
CC   responsible for genetic disorders or other traits to assess biodiversity
CC   and to produce other types of data and products dependent on DNA and
CC   amino acid sequences. ABG00010-ABG30377 represent novel human diagnostic
CC   amino acid sequences of the invention. Note: The sequence data for this
CC   patent did not appear in the printed specification, but was obtained in
CC   electronic format directly from WIPO at
CC   ftp.wipo.int/pub/published_pct_sequences
XX
SQ   Sequence 160 AA;

  Query Match             36.4%;  Score 200.5;  DB 4;  Length 160;
  Best Local Similarity   73.0%;  Pred. No. 1.9e-14;
  Matches   46;  Conservative    2;  Mismatches   12;  Indels    3;  Gaps    2;

Qy          9 LKKKAQYEANKVKLW-GLNTEYDNLLFVINSIKQEIVNRVQNPRDERGPNMGQKLEILIK 67
              | |   |: :   ||     | ||||||||||||||||||||||||||||||||||||||
Db          8 LTKNHGYDHS--SLWRSTALEPYNLLFVINSIKQEIVNRVQNPRDERGPNMGQKLEILIK 65

Qy         68 DTL 70
              |||
Db         66 DTL 68


RESULT 5
ABU11886
ID   ABU11886 standard; protein; 1124 AA.
XX
AC   ABU11886;
XX
DT   13-FEB-2003  (first entry)
XX
DE   Human ABCA1 interacting polypeptide KIAA0528.
XX
KW   High throughput screening assay; ATP-binding cassette transporter 1;
KW   ABCA1; cholesterol efflux regulating protein; CERP; AIP; human;
KW   ABCA1-interacting protein; plasma cholesterol; plasma phospholipid;
KW   lipid metabolism disorder; cardiovascular disease; CVD; CAD;
KW   coronary artery disease; cerebrovascular disease; coronary restenosis;
KW   peripheral vascular disease; anti-lipaemic.
XX
OS   Homo sapiens.
XX
PN   WO200284301-A2.
XX
PD   24-OCT-2002.
XX
PF   12-APR-2002; 2002WO-CA000489.
XX
PR   12-APR-2001; 2001US-0283424P.
XX
PA   (XENO-) XENON GENETICS INC.
PA   (UYBR-) UNIV BRITISH COLUMBIA.
XX
PI   Zhang L,  Hayden MR,  Newman SM;
XX
DR   WPI; 2003-093046/08.
DR   N-PSDB; ABX55992.
XX
PT   Identifying ABCA1 (cholesterol efflux regulating protein) modulators for
PT   modulating cholesterol or phospholipid levels in animals, by determining
PT   ABCA1 activity in the presence of an agent-contacted ABCA1-interacting
PT   protein.
XX
PS   Disclosure; Page 59-63; 92pp; English.
XX
CC   The present invention relates to high throughput screening assays for
CC   agents capable of modulating the activity of ATP-binding cassette
CC   transporter 1 (ABCA1, also called cholesterol efflux regulating protein
CC   (CERP)). The method comprises determining the activity of an ABCA1
CC   polypeptide in the presence of an ABCA1-interacting protein (AIP) that
CC   has been contacted with an agent. The method is useful for identifying
CC   modulators of ABCA1 biological activity, particularly for identifying
CC   agents that modulate cholesterol or phospholipid levels in an animal. The
CC   method is useful for treating disorders of lipid metabolism, especially
CC   for reducing elevated plasma phospholipid or cholesterol levels in
CC   patients, or for preventing elevated phospholipid or cholesterol levels
CC   in a patient at risk of developing disorders such as cardiovascular
CC   disease (CVD), coronary artery disease (CAD), cerebrovascular disease,
CC   coronary restenosis, and peripheral vascular disease. The present
CC   sequence represents a human AIP polypeptide
XX
SQ   Sequence 1124 AA;

  Query Match             14.9%;  Score 82;  DB 6;  Length 1124;
```

```
  Best Local Similarity   28.4%;  Pred. No. 7.5;
  Matches   29;  Conservative   22;  Mismatches   27;  Indels   24;  Gaps    6;

Qy          5 RACRLKKKAQYEANKVKLWGLNTEYDNLL-PVINSIKQEIVNRVQNPRDERGPN--MGQK 61
              | |||||||| |||        |   ||| |:   :  :::|:::     :| |   | :
Db        590 RLCRLKKKAQAEANA-------TAISNLLPFMEYEVHTQLMNKLK----LKGMNALFGLR 638

Qy         62 LEILIKDTLGLPVAGQTSEFVNQVLEKTAEGNPTGGLVGLRI 103
              ::| : : : : :|  |  ::  :        || |  |::|
Db        639 IQITVGENMLMGLASATGVYLAAL--------PTPG--GIQI 670


RESULT 6
ABU19277
ID   ABU19277 standard; protein; 281 AA.
XX
AC   ABU19277;
XX
DT   19-JUN-2003  (first entry)
XX
DE   Protein encoded by Prokaryotic essential gene #4804.
XX
KW   Antisense; prokaryotic essential gene; cell proliferation; drug design.
XX
OS   Borrelia burgdorferi.
XX
PN   WO200277183-A2.
XX
PD   03-OCT-2002.
XX
PF   21-MAR-2002; 2002WO-US009107.
XX
PR   21-MAR-2001; 2001US-00815242.
PR   06-SEP-2001; 2001US-00948993.
PR   25-OCT-2001; 2001US-0342923P.
PR   08-FEB-2002; 2002US-00072851.
PR   06-MAR-2002; 2002US-0362699P.
XX
PA   (ELIT-) ELITRA PHARM INC.
XX
PI   Wang L,  Zamudio C,  Malone C,  Haselbeck R,  Ohlsen KL,  Zyskind JW;
PI   Wall D,  Trawick JD,  Carr GJ,  Yamamoto R,  Forsyth RA,  Xu HH;
XX
DR   WPI; 2003-029926/02.
DR   N-PSDB; ACA23147.
XX
PT   New antisense nucleic acids, useful for identifying proteins or screening
PT   for homologous nucleic acids required for cellular proliferation to
PT   isolate candidate molecules for rational drug discovery programs.
XX
PS   Claim 25; SEQ ID NO 47201; 1766pp; English.
XX
CC   The invention relates to an isolated nucleic acid comprising any one of
CC   the 6213 antisense sequences given in the specification where expression
CC   of the nucleic acid inhibits proliferation of a cell. Also included are:
CC   (1) a vector comprising a promoter operably linked to the nucleic acid
CC   encoding a polypeptide whose expression is inhibited by the antisense
CC   nucleic acid; (2) a host cell containing the vector; (3) an isolated
CC   polypeptide or its fragment whose expression is inhibited by the
CC   antisense nucleic acid; (4) an antibody capable of specifically binding
CC   the polypeptide;(5) producing the polypeptide; (6) inhibiting cellular
CC   proliferation or the activity of a gene in an operon required for
CC   proliferation; (7) identifying a compound that influences the activity of
CC   the gene product or that has an activity against a biological pathway
CC   required for proliferation, or that inhibits cellular proliferation; (8)
CC   identifying a gene required for cellular proliferation or the biological
CC   pathway in which a proliferation-required gene or its gene product lies
CC   or a gene on which the test compound that inhibits proliferation of an
CC   organism acts; (9) manufacturing an antibiotic; (10) profiling a
CC   compound's activity; (11) a culture comprising strains in which the gene
CC   product is overexpressed or underexpressed; (12) determining the extent
CC   to which each of the strains is present in a culture or collection of
CC   strains; or (13) identifying the target of a compound that inhibits the
CC   proliferation of an organism. The antisense nucleic acids are useful for
CC   identifying proteins or screening for homologous nucleic acids required
CC   for cellular proliferation to isolate candidate molecules for rational
CC   drug discovery programs, or for screening homologous nucleic acids
CC   required for proliferation in cells other than S. aureus, S. typhimurium,
CC   K. pneumoniae or P. aeruginosa. The present sequence is encoded by one of
CC   the target prokaryotic essential genes. Note: The sequence data for this
CC   patent did not form part of the printed specification, but was obtained
CC   in electronic format directly from WIPO at
CC   ftp.wipo.int/pub/published_pct_sequences
XX
SQ   Sequence 281 AA;

  Query Match             14.2%;  Score 78.5;  DB 6;  Length 281;
  Best Local Similarity   31.6%;  Pred. No. 3.2;
  Matches   24;  Conservative   14;  Mismatches   23;  Indels   15;  Gaps    4;

Qy         21 KLWGLNTEYDNLLFVIN-SIKQEIVNRVQNPRDER----GPNMGQKLEILIKDT---LGL 72
              |||| |       ::|| ||:|:|:  :    :|:    || :|   |||:| |
Db         31 KLWGQN-------YLINESIRQKIIESLDIKENEKIWEIGPGLGAMTEILLKKTNLLTAF 83

Qy         73 PVAGQTSEFVNQVLEK 88
              :  : || :|:   |
Db         84 EIDLKYSEILNEKFGK 99


RESULT 7
ABU01541
ID   ABU01541 standard; protein; 345 AA.
XX
AC   ABU01541;
XX
DT   23-OCT-2003  (revised)
DT   11-FEB-2003  (first entry)
XX
DE   S. pneumoniae type 4 strain protein from coding region #1117.
XX
KW   Bacterial meningitis; pneumonia; sepsis; otitis media; ear infection;
KW   antiinflammatory; antibacterial; immunostimulant; auditory; respiratory;
KW   gene therapy; vaccine.
XX
OS   Streptococcus pneumoniae; type 4 strain.
XX
PN   WO200277021-A2.
XX
PD   03-OCT-2002.
XX
PF   27-MAR-2002; 2002WO-IB002163.
XX
PR   27-MAR-2001; 2001GB-00007658.
XX
PA   (CHIR-) CHIRON SPA.
PA   (GENO-) INST GENOMIC RES.
XX
PI   Masignani V,  Tettelin H,  Fraser C;
XX
DR   WPI; 2003-040579/03.
DR   N-PSDB; ABX06829.
XX
PT   New proteins and nucleic acid molecules from Streptococcus pneumoniae,
PT   useful as medicaments for treating or preventing a disease or infection
PT   due to streptococcus bacteria, such as pneumonia, sepsis, otitis media or
PT   ear infection.
XX
PS   Claim 1; SEQ ID NO 2234; 56pp; English.
XX
CC   The invention relates to a protein comprising or having at least 50%
CC   identity to any of the 2469 amino acid sequences, identified in the
CC   specification (available on a computer readable format), or its fragment,
CC   expressed from 2469 of 2489 identified DNA coding regions from the
CC   Streptococcus pneumoniae type 4 strain genomic sequence appearing as
CC   ABS56454. Also included are an antibody which binds one of the proteins,
CC   treating a patient by administering the protein, DNA or antibody (in a
```

CC composition), a kit comprising first and second primers, which are the
CC nucleic acid cited above or fragments between nucleotides 8-100 of a
CC sequence not defined in the specification, for amplifying a target
CC sequence contained within a Streptococcus nucleic acid sequence, where
CC the first primer is substantially complementary to the target sequence
CC and the second primer is substantially complementary to the complement of
CC the target sequence, and where the parts of the primers having
CC substantial complementarity define the termini of the target sequence to
CC be amplified, assay comprising contacting a test compound with the
CC protein, and determining whether the test compound binds to the protein
CC and a Streptococcus pneumoniae bacterium, where one or more genes
CC encoding the proteins has been rendered inactive. The proteins, nucleic
CC acid molecules, antibody and compositions are useful as medicaments for
CC treating or preventing a disease or infection due to streptococcus
CC bacteria, particularly S. pneumoniae, such as pneumonia, sepsis, otitis
CC media or ear infection. They are also useful in developing vaccines,
CC diagnostics and antibiotics. The methods are useful for identifying
CC immunodominant proteins. The present sequence is one of the 2469 proteins
CC expressed by the identified coding regions from the genomic sequence.
CC Note: The sequence data for this patent did not form part of the printed
CC specification, but was obtained in electronic format directly from WIPO
CC at ftp.wipo.int/pub/published_pct_sequences. (Updated on 23-OCT-2003 to
CC standardise OS field)
XX
SQ Sequence 345 AA;

Query Match 13.7%; Score 75.5; DB 6; Length 345;
Best Local Similarity 27.6%; Pred. No. 9.1;
Matches 29; Conservative 17; Mismatches 20; Indels 39; Gaps 6;

```
Qy          9 LKKKAQYEANKVK--------LWGLNTEYDNLLFVINSIKQEIVNRVQNPRDERGPNM-- 58
              :  | :| |:::          || ::| || :          ||::      | |  ::
Db        193 MNAKMKYFASQLNQFYKDHRCLWEIDTSYDGI---------EIID-----ADNRDQSVLS 238

Qy         59 ----GQKLEILI----------KD-TLGLPVAGQTSEFVNQVLEK 88
                  |:| |:|:          || |:||||||   |  |  ||:
Db        239 FIRKGKKGEMLVCIFNMVPVERKDFTIGLPVAGIYEEVWNTELEE 283
```

RESULT 8
ADM92163
ID ADM92163 standard; protein; 642 AA.
XX
AC ADM92163;
XX
DT 03-JUN-2004 (first entry)
XX
DE S pneumoniae antigenic protein sequence SeqID360.
XX
KW antibacterial; gene therapy; Streptococcus pneumoniae infection;
KW antigenic.
XX
OS Streptococcus pneumoniae.
XX
PN WO2004020609-A2.
XX
PD 11-MAR-2004.
XX
PF 02-SEP-2003; 2003WO-US027401.
XX
PR 30-AUG-2002; 2002US-0407082P.
XX
PA (TUFT) UNIV TUFTS.
XX
PI Camilli A, Hava DL;
XX
DR WPI; 2004-239189/22.
DR N-PSDB; ADM91926.
XX
PT New Streptococcus pneumoniae nucleic acid molecules, useful for
PT diagnosing, treating and preventing active infections of Streptococcus
PT pneumoniae.
XX
PS Claim 27; SEQ ID NO 360; 123pp; English.
XX
CC This invention relates to novel isolated Streptococcus pneumoniae nucleic
CC acid molecules and the antigenic polypeptides encoded by them. The
CC invention may be useful for the production of compounds with an
CC antibacterial activity or for gene therapy. The nucleic acid molecules,
CC compositions and methods disclosed are useful for treating Streptococcus
CC pneumoniae infection. The present sequence is that of an S pneumoniae
CC protein of the invention.
XX
SQ Sequence 642 AA;

Query Match 13.7%; Score 75.5; DB 8; Length 642;
Best Local Similarity 27.6%; Pred. No. 20;
Matches 29; Conservative 17; Mismatches 20; Indels 39; Gaps 6;

```
Qy          9 LKKKAQYEANKVK--------LWGLNTEYDNLLFVINSIKQEIVNRVQNPRDERGPNM-- 58
              :  | :| |:::          || ::| || :          ||::      | |  ::
Db        490 MNAKMKYFASQLNQFYKDHRCLWEIDTSYDGI---------EIID-----ADNRDQSVLS 535

Qy         59 ----GQKLEILI----------KD-TLGLPVAGQTSEFVNQVLEK 88
                  |:| |:|:          || |:||||||   |  |  ||:
Db        536 FIRKGKKGEMLVCIFNMVPVERKDFTIGLPVAGIYEEVWNTELEE 580
```

RESULT 9
ADS44237
ID ADS44237 standard; protein; 1175 AA.
XX
AC ADS44237;
XX
DT 02-DEC-2004 (first entry)
XX
DE Bacterial polypeptide #22667.
XX
KW Recombinant DNA construct; transformed plant; improved plant property;
KW cold tolerance; heat tolerance; drought tolerance; herbicide; osmosis;
KW pathogen tolerance; pest tolerance; plant disease resistance;
KW cell cycle pathway modification; plant growth regulator;
KW homologous recombination; seed oil yield; protein yield; carbohydrate;
KW nitrogen; phosphorus; photosynthesis; lignin; galactomannan;
KW bacterial polypeptide.
XX
OS Bacteria.
XX
PN US2003233675-A1.
XX
PD 18-DEC-2003.
XX
PF 20-FEB-2003; 2003US-00369493.
XX
PR 21-FEB-2002; 2002US-0360039P.
XX
PA (CAOY/) CAO Y.
PA (HINK/) HINKLE G J.
PA (SLAT/) SLATER S C.
PA (CHEN/) CHEN X.
PA (GOLD/) GOLDMAN B S.
XX
PI Cao Y, Hinkle GJ, Slater SC, Chen X, Goldman BS;
XX
DR WPI; 2004-061375/06.
XX
PT New recombinant DNA construct comprising a promoter positioned to provide
PT for expression of a polynucleotide encoding a polypeptide from a
PT microbial source, useful for producing plants with improved properties.
XX
PS Claim 1; SEQ ID NO 22667; 122pp; English.
XX
CC The invention relates to a recombinant DNA construct comprising a
CC promoter functional in a plant cell, where the promoter is positioned to
CC provide for expression of a polynucleotide encoding a polypeptide from a
CC microbial source. The invention also relates to a transformed plant

CC comprising the recombinant DNA construct and a method of producing a
CC transformed plant having an improved property. The plant is a crop plant
CC such as maize or soybean. The method of producing a transformed plant
CC having an improved property comprises transforming a plant with the
CC recombinant DNA construct and growing the transformed plant, where the
CC polynucleotide or polypeptide is useful for improving plant properties.
CC The recombinant DNA construct is useful for producing plants with
CC improved plant properties, e.g. improved cold, heat or drought tolerance,
CC tolerance to herbicides, extreme osmotic conditions, pathogens or pests,
CC increased resistance to plant disease, better growth rate by modification
CC of the cell cycle pathway with plant growth regulators, increased rate of
CC homologous recombination, modified seed oil or protein yield and/or
CC content, improved yield by modification of carbohydrate, nitrogen or
CC phosphorus use and/or uptake, by modification of photosynthesis or by
CC providing improved plant growth and development under at least one stress
CC condition, improved lignin production or improved galactomannan
CC production. This sequence represents a bacterial polypeptide used in the
CC scope of the invention. Note: The sequence data for this patent did not
CC form part of the printed specification but was obtained in electronic
CC format from USPTO at seqdata.uspto.gov/sequence.html.
XX
SQ Sequence 1175 AA;

Query Match 13.7%; Score 75.5; DB 8; Length 1175;
Best Local Similarity 25.0%; Pred. No. 44;
Matches 23; Conservative 21; Mismatches 41; Indels 7; Gaps 2;

```
Qy          2 LAFRACRLKKKAQYEANKVKLWGLNTEYDNLLFVINSIKQEIVNRVQNPRDERGPNMGQK 61
              :   :||   : : :  : : :     | |:||:|:::|:   :| |: | ||   : |:
Db         38 VCLNSCRFYNETRSQLREKRFFVGXNLYINILFLISNVKRIRASRNQDNRPERQSRL-QR 96

Qy         62 LEILIKDTLGLPVAGQTSEFVNQVLEKTAEGN 93
                ||:   |       :  | || :| | |
Db         97 SSSLIEQVRG------NEDGENDVLNQTRETN 122
```

RESULT 10
ADS30935
ID ADS30935 standard; protein; 1124 AA.
XX
AC ADS30935;
XX
DT 02-DEC-2004 (first entry)
XX
DE Bacterial polypeptide #19968.
XX
KW Recombinant DNA construct; transformed plant; improved plant property;
KW cold tolerance; heat tolerance; drought tolerance; herbicide; osmosis;
KW pathogen tolerance; pest tolerance; plant disease resistance;
KW cell cycle pathway modification; plant growth regulator;
KW homologous recombination; seed oil yield; protein yield; carbohydrate;
KW nitrogen; phosphorus; photosynthesis; lignin; galactomannan;
KW bacterial polypeptide.
XX
OS Bacteria.
XX
PN US2003233675-A1.
XX
PD 18-DEC-2003.
XX
PF 20-FEB-2003; 2003US-00369493.
XX
PR 21-FEB-2002; 2002US-0360039P.
XX
PA (CAOY/) CAO Y.
PA (HINK/) HINKLE G J.
PA (SLAT/) SLATER S C.
PA (CHEN/) CHEN X.
PA (GOLD/) GOLDMAN B S.
XX
PI Cao Y, Hinkle GJ, Slater SC, Chen X, Goldman BS;
XX
DR WPI; 2004-061375/06.
XX
PT New recombinant DNA construct comprising a promoter positioned to provide
PT for expression of a polynucleotide encoding a polypeptide from a
PT microbial source, useful for producing plants with improved properties.
XX
PS Claim 1; SEQ ID NO 19968; 122pp; English.
XX
CC The invention relates to a recombinant DNA construct comprising a
CC promoter functional in a plant cell, where the promoter is positioned to
CC provide for expression of a polynucleotide encoding a polypeptide from a
CC microbial source. The invention also relates to a transformed plant
CC comprising the recombinant DNA construct and a method of producing a
CC transformed plant having an improved property. The plant is a crop plant
CC such as maize or soybean. The method of producing a transformed plant
CC having an improved property comprises transforming a plant with the
CC recombinant DNA construct and growing the transformed plant, where the
CC polynucleotide or polypeptide is useful for improving plant properties.
CC The recombinant DNA construct is useful for producing plants with
CC improved plant properties, e.g. improved cold, heat or drought tolerance,
CC tolerance to herbicides, extreme osmotic conditions, pathogens or pests,
CC increased resistance to plant disease, better growth rate by modification
CC of the cell cycle pathway with plant growth regulators, increased rate of
CC homologous recombination, modified seed oil or protein yield and/or
CC content, improved yield by modification of carbohydrate, nitrogen or
CC phosphorus use and/or uptake, by modification of photosynthesis or by
CC providing improved plant growth and development under at least one stress
CC condition, improved lignin production or improved galactomannan
CC production. This sequence represents a bacterial polypeptide used in the
CC scope of the invention. Note: The sequence data for this patent did not
CC form part of the printed specification but was obtained in electronic
CC format from USPTO at seqdata.uspto.gov/sequence.html.
XX
SQ Sequence 1124 AA;

Query Match 13.5%; Score 74.5; DB 8; Length 1124;
Best Local Similarity 28.6%; Pred. No. 54;
Matches 26; Conservative 11; Mismatches 37; Indels 17; Gaps 4;

```
Qy         24 GLNTEYDNLLFVINSIKQEIVNRVQNPRDERGPNMGQKLEILIKDTLGLPVA-------- 75
              |: |   |: | :| :  ::|    |||  |   :  |          | |:|
Db        316 GVETGGSNIQFAVNPLNGDVVVIEMNPRVSRSSALSSK-------ATGFPIAKMAAKLAV 368

Qy         76 GQT-SEFVNQVLEKT-AEGNPTGGLVGLRIP 104
              | |  |  | : :|| |   ||   | :||
Db        369 GYTLDEIKNDITKKTPASFEPTIDYVVTKIP 399
```

RESULT 11
ABU43663
ID ABU43663 standard; protein; 625 AA.
XX
AC ABU43663;
XX
DT 19-JUN-2003 (first entry)
XX
DE Protein encoded by Prokaryotic essential gene #29190.
XX
KW Antisense; prokaryotic essential gene; cell proliferation; drug design.
XX
OS Staphylococcus haemolyticus.
XX
PN WO200277183-A2.
XX
PD 03-OCT-2002.
XX
PF 21-MAR-2002; 2002WO-US009107.
XX
PR 21-MAR-2001; 2001US-00815242.
PR 06-SEP-2001; 2001US-00948993.
PR 25-OCT-2001; 2001US-0342923P.
PR 08-FEB-2002; 2002US-00072851.
PR 06-MAR-2002; 2002US-0362699P.
XX

PA (ELIT-) ELITRA PHARM INC.
XX
PI Wang L, Zamudio C, Malone C, Haselbeck R, Ohlsen KL, Zyskind JW;
PI Wall D, Trawick JD, Carr GJ, Yamamoto R, Forsyth RA, Xu HH;
XX
DR WPI; 2003-029926/02.
DR N-PSDB; ACA47533.
XX
PT New antisense nucleic acids, useful for identifying proteins or screening
PT for homologous nucleic acids required for cellular proliferation to
PT isolate candidate molecules for rational drug discovery programs.
XX
PS Claim 25; SEQ ID NO 71587; 1766pp; English.
XX
CC The invention relates to an isolated nucleic acid comprising any one of
CC the 6213 antisense sequences given in the specification where expression
CC of the nucleic acid inhibits proliferation of a cell. Also included are:
CC (1) a vector comprising a promoter operably linked to the nucleic acid
CC encoding a polypeptide whose expression is inhibited by the antisense
CC nucleic acid; (2) a host cell containing the vector; (3) an isolated
CC polypeptide or its fragment whose expression is inhibited by the
CC antisense nucleic acid; (4) an antibody capable of specifically binding
CC the polypeptide;(5) producing the polypeptide; (6) inhibiting cellular
CC proliferation or the activity of a gene in an operon required for
CC proliferation; (7) identifying a compound that influences the activity of
CC the gene product or that has an activity against a biological pathway
CC required for proliferation, or that inhibits cellular proliferation; (8)
CC identifying a gene required for cellular proliferation or the biological
CC pathway in which a proliferation-required gene or its gene product lies
CC or a gene on which the test compound that inhibits proliferation of an
CC organism acts; (9) manufacturing an antibiotic; (10) profiling a
CC compound's activity; (11) a culture comprising strains in which the gene
CC product is overexpressed or underexpressed; (12) determining the extent
CC to which each of the strains is present in a culture or collection of
CC strains; or (13) identifying the target of a compound that inhibits the
CC proliferation of an organism. The antisense nucleic acids are useful for
CC identifying proteins or screening for homologous nucleic acids required
CC for cellular proliferation to isolate candidate molecules for rational
CC drug discovery programs, or for screening homologous nucleic acids
CC required for proliferation in cells other than S. aureus, S. typhimurium,
CC K. pneumoniae or P. aeruginosa. The present sequence is encoded by one of
CC the target prokaryotic essential genes. Note: The sequence data for this
CC patent did not form part of the printed specification, but was obtained
CC in electronic format directly from WIPO at
CC ftp.wipo.int/pub/published_pct_sequences
XX
SQ Sequence 625 AA;

Query Match 13.3%; Score 73.5; DB 6; Length 625;
Best Local Similarity 21.8%; Pred. No. 33;
Matches 22; Conservative 17; Mismatches 31; Indels 31; Gaps 2;

Qy 8 RLKKKAQYEANKVKLWGLNTEYDNLLF----------VINSIKQEI 43
Db 130 RLQTRLSYTESQIKFYKTKTVYTDLLFLYLEKAFLSQNFFDIPMMHSQLDDILNSMFQYL 189

Qy 44 VNRVQNPRDERGPNMGQKLEILIKDTLGLPVAGQTSEFVNQ 84
Db 190 PNFFQNETSEDNMYLAQRIMFQIDDML-------TKDMLNE 223

RESULT 12
ADS30054
ID ADS30054 standard; protein; 1104 AA.
XX
AC ADS30054;
XX
DT 02-DEC-2004 (first entry)
XX
DE Bacterial polypeptide #19087.
XX
KW Recombinant DNA construct; transformed plant; improved plant property;
KW cold tolerance; heat tolerance; drought tolerance; herbicide; osmosis;
KW pathogen tolerance; pest tolerance; plant disease resistance;
KW cell cycle pathway modification; plant growth regulator;
KW homologous recombination; seed oil yield; protein yield; carbohydrate;
KW nitrogen; phosphorus; photosynthesis; lignin; galactomannan;
KW bacterial polypeptide.
XX
OS Bacteria.
XX
PN US2003233675-A1.
XX
PD 18-DEC-2003.
XX
PF 20-FEB-2003; 2003US-00369493.
XX
PR 21-FEB-2002; 2002US-0360039P.
XX
PA (CAOY/) CAO Y.
PA (HINK/) HINKLE G J.
PA (SLAT/) SLATER S C.
PA (CHEN/) CHEN X.
PA (GOLD/) GOLDMAN B S.
XX
PI Cao Y, Hinkle GJ, Slater SC, Chen X, Goldman BS;
XX
DR WPI; 2004-061375/06.
XX
PT New recombinant DNA construct comprising a promoter positioned to provide
PT for expression of a polynucleotide encoding a polypeptide from a
PT microbial source, useful for producing plants with improved properties.
XX
PS Claim 1; SEQ ID NO 19087; 122pp; English.
XX
CC The invention relates to a recombinant DNA construct comprising a
CC promoter functional in a plant cell, where the promoter is positioned to
CC provide for expression of a polynucleotide encoding a polypeptide from a
CC microbial source. The invention also relates to a transformed plant
CC comprising the recombinant DNA construct and a method of producing a
CC transformed plant having an improved property. The plant is a crop plant
CC such as maize or soybean. The method of producing a transformed plant
CC having an improved property comprises transforming a plant with the
CC recombinant DNA construct and growing the transformed plant, where the
CC polynucleotide or polypeptide is useful for improving plant properties.
CC The recombinant DNA construct is useful for producing plants with
CC improved plant properties, e.g. improved cold, heat or drought tolerance,
CC tolerance to herbicides, extreme osmotic conditions, pathogens or pests,
CC increased resistance to plant disease, better growth rate by modification
CC of the cell cycle pathway with plant growth regulators, increased rate of
CC homologous recombination, modified seed oil or protein yield and/or
CC content, improved yield by modification of carbohydrate, nitrogen or
CC phosphorus use and/or uptake, by modification of photosynthesis or by
CC providing improved plant growth and development under at least one stress
CC condition, improved lignin production or improved galactomannan
CC production. This sequence represents a bacterial polypeptide used in the
CC scope of the invention. Note: The sequence data for this patent did not
CC form part of the printed specification but was obtained in electronic
CC format from USPTO at seqdata.uspto.gov/sequence.html.
XX
SQ Sequence 1104 AA;

Query Match 13.3%; Score 73.5; DB 8; Length 1104;
Best Local Similarity 27.5%; Pred. No. 68;
Matches 25; Conservative 12; Mismatches 37; Indels 17; Gaps 4;

Qy 24 GLNTEYDNLLFVINSIKQEIVNRVQNPRDERGPNMGQKLEILIKDTLGLPVA-------- 75
Db 281 GVETGGSNIQFAVNPVNGDVVVIEMNPRVSRSSALSSK-------ATGFPIAKMAAKLAV 333

Qy 76 GQT-SEFVNQVLEKT-AEGNPTGGLVGLRIP 104
Db 334 GYTLDEIRNDITKKTPASFEPTIDYVVTKVP 364

RESULT 13

```
AAW18226
ID   AAW18226 standard; protein; 1495 AA.
XX
AC   AAW18226;
XX
DT   24-SEP-1997  (first entry)
XX
DE   Transcriptional co-repressor SMRT.
XX
KW   Silencing mediator for retinoic acid and thyroid hormone receptor; SMRT;
KW   transcriptional co-repressor.
XX
OS   Homo sapiens.
XX
FH   Key             Location/Qualifiers
FT   Region          1. .160
FT                   /label= N-terminal_region
FT                   /note= "proline-rich domain"
FT   Region          773. .790
FT                   /label= ERDR_region
FT   Region          812. .827
FT                   /label= SG_region
FT   Region          1061. .1132
FT                   /label= glutamine-rich region
FT   Region          1201. .1495
FT                   /label= C-terminal_region
FT   Peptide         1330. .1376
FT                   /note= "alternatively spliced insert not present in the
FT                   original two-hybrid clone"
XX
PN   WO9709418-A1.
XX
PD   13-MAR-1997.
XX
PF   24-JUL-1996;   96WO-US012371.
XX
PR   01-SEP-1995;   95US-00522726.
XX
PA   (SALK ) SALK INST BIOLOGICAL STUDIES.
XX
PI   Evans RM,  Chen JD;
XX
DR   WPI; 1997-192894/17.
XX
PT   New co-suppressor of steroid-thyroid hormone receptor activity - also
PT   methods for identifying compounds that relieve its suppressant effect
PT   and/or activate receptors.
XX
PS   Claim 2; Page 40-45; 71pp; English.
XX
CC   A novel receptor interacting factor (AAW18226) is designated SMRT, i.e.
CC   silencing mediator (co-repressor) for retinoic acid receptor (RAR) and
CC   thyroid hormone receptor (TR). Its association with RAR and TR both in
CC   solution and on DNA response elements is destabilised by ligand. The
CC   interaction of SMRT with mutant receptors correlates with the
CC   transcriptional silencing activities of receptors. In vivo, SMRT
CC   functions as a potent co-repressor. A GAL4 DNA binding domain fusion of
CC   SMRT behaves as a frank repressor of a GAL4-dependent reporter. These
CC   data identify a novel class of cofactor which is believed to represent an
CC   important mediator of hormone action. Full-length cDNA for SMRT has been
CC   isolated from a HeLa library in a two-hybrid screen using a GAL4 DNA
CC   binding domain/RXR fusion protein as bait
XX
SQ   Sequence 1495 AA;

  Query Match             13.3%;  Score 73.5;  DB 2;  Length 1495;
  Best Local Similarity   30.6%;  Pred. No. 1e+02;
  Matches   22;  Conservative   14;  Mismatches   17;  Indels   19;  Gaps    3;

Qy         37 NSIKQEIVNRVQNPRDERGPNMGQKLEILIKDTLGLPVAGQTSEFVNQVLEKTAEGNPTG 96
              : :||| :    :|| : ||          ::||:| | : |     ||  || |: |
Db        143 SGVKQEQL----SPRGQAGP----------PESLGVPTAQEAS-----VLRGTALGSVPG 183

Qy         97 GLVGLRIPTSKV 108
              | :   ||:::|
Db        184 GSITKGIPSTRV 195


RESULT 14
ADN04305
ID   ADN04305 standard; protein; 2000 AA.
XX
AC   ADN04305;
XX
DT   01-JUL-2004  (first entry)
XX
DE   Antipsoriatic protein sequence #347.
XX
KW   antipsoriatic; gene therapy; psoriasis; diagnosis.
XX
OS   Homo sapiens.
XX
PN   WO2004028479-A2.
XX
PD   08-APR-2004.
XX
PF   25-SEP-2003; 2003WO-US030907.
XX
PR   25-SEP-2002; 2002US-0414006P.
XX
PA   (GETH ) GENENTECH INC.
XX
PI   Bodary S,  Clark H,  Jackman J,  Schoenfeld J,  Williams PM,  Wood WI;
PI   Wu TD;
XX
DR   WPI; 2004-305105/28.
DR   N-PSDB; ADN04304.
XX
PT   New PRO nucleic acid or polypeptide, useful for preparing a
PT   pharmaceutical composition for diagnosing or treating psoriasis in a
PT   mammal.
XX
PS   Claim 9; SEQ ID NO 699; 3069pp; English.
XX
CC   The invention relates to novel polynucleotide and polypeptides for
CC   treating psoriasis or a sequence having at least 80% identity to the
CC   above sequences. The nucleic acid is useful for preparing a composition
CC   for diagnosing or treating psoriasis in a mammal. This sequence
CC   corresponds to one of the polypeptides of the invention.
XX
SQ   Sequence 2000 AA;

  Query Match             13.3%;  Score 73.5;  DB 8;  Length 2000;
  Best Local Similarity   30.6%;  Pred. No. 1.5e+02;
  Matches   22;  Conservative   14;  Mismatches   17;  Indels   19;  Gaps    3;

Qy         37 NSIKQEIVNRVQNPRDERGPNMGQKLEILIKDTLGLPVAGQTSEFVNQVLEKTAEGNPTG 96
              : :||| :    :|| : ||          ::||:| | : |     ||  || |: |
Db       1165 SGVKQEQL----SPRGQAGP----------PESLGVPTAQEAS-----VLRGTALGSVPG 1205

Qy         97 GLVGLRIPTSKV 108
              | :   ||:::|
Db       1206 GSITKGIPSTRV 1217


RESULT 15
ABU61812
ID   ABU61812 standard; protein; 2507 AA.
XX
AC   ABU61812;
XX
DT   12-AUG-2003  (first entry)
XX
DE   Human nuclear receptor corepressor SMRTe.
XX
KW   Human; SMRTe; nuclear receptor corepressor; gene therapy; tissue typing;
```

```
KW   cancer.
XX
OS   Homo sapiens.
XX
FH   Key             Location/Qualifiers
FT   Domain          166. .429
FT                   /label= SNC_domain
FT   Region          166. .206
FT                   /label= Amphipathic_helix
FT                   /note= "Residues 172, 179, 186, 194 and 200 make up the
FT                   hydrophobic heptad repeat"
FT   Domain          430. .486
FT                   /label= SANT_A_domain
FT   Domain          613. .669
FT                   /label= SANT_B_domain
XX
PN   US2003027137-A1.
XX
PD   06-FEB-2003.
XX
PF   27-MAR-2001; 2001US-00819104.
XX
PR   29-MAR-2000; 2000US-0193138P.
XX
PA   (CHEN/) CHEN J D.
XX
PI   Chen JD;
XX
DR   WPI; 2003-466139/44.
DR   N-PSDB; ACA62249, ACA62250.
XX
PT   New SMRTe proteins and nucleic acids, useful in gene therapy, predictive
PT   medicine, therapeutic or prophylactic treatment, chromosome mapping,
PT   tissue typing and in forensic biology.
XX
PS   Claim 9; Fig 1; 90pp; English.
XX
CC   The invention relates to an isolated SMRTe nucleic acid molecule. The
CC   nucleic acids are useful in gene therapy, as hybridisation probes for
CC   identifying SMRTe-encoding nucleic acid molecules and as primers for
CC   amplifying of SMRTe nucleic acid molecules. The polypeptides are useful
CC   as immunogens to raise anti-SMRTe antibodies. The SMRTe molecules are
CC   useful as targets for discovering and developing modulating agents to
CC   regulate a variety of cellular processes, in screening assays, in
CC   predictive medicine, in therapeutic or prophylactic treatment, in
CC   chromosome mapping, tissue typing and in forensic identification of a
CC   biological sample. Modulators of SMRTe are useful for treating or
CC   preventing a condition associated with aberrant SMRTe protein or nucleic
CC   acid expression or activity, such as cancer. The present sequence
CC   represents the amino acid sequence of the human nuclear receptor
CC   corepressor SMRTe
XX
SQ   Sequence 2507 AA;

  Query Match             13.3%;  Score 73.5;  DB 6;  Length 2507;
  Best Local Similarity   30.6%;  Pred. No. 2e+02;
  Matches   22;  Conservative   14;  Mismatches   17;  Indels   19;  Gaps    3;

Qy         37 NSIKQEIVNRVQNPRDERGPNMGQKLEILIKDTLGLPVAGQTSEFVNQVLEKTAEGNPTG 96
              : :||| :    :|| : ||           ::||:| | : |     ||  || |:  |
Db       1155 SGVKQEQL----SPRGQAGP----------PESLGVPTAQEAS-----VLRGTALGSVPG 1195

Qy         97 GLVGLRIPTSKV 108
              | :   ||:::|
Db       1196 GSITKGIPSTRV 1207


Search completed: February 28, 2006, 08:45:11
Job time : 39.9925 secs
```

This Page Blank (uspto)

GenCore version 5.1.7

OM protein - protein search, using sw model

Run on: February 28, 2006, 08:53:26 ; Search time 9.29323 Seconds
(without alignments)
960.804 Million cell updates/sec

Title: US-10-717-665A-44_COPY_532_639
Perfect score: 551
Sequence: 1 KLAFRACRLKKKAQYEANKV.....TAEGNPTGGLVGLRIPTSKV 108

Scoring table: BLOSUM62
Gapop 10.0 , Gapext 0.5

Searched: 572060 seqs, 82675679 residues

Total number of hits satisfying chosen parameters: 572060

Minimum DB seq length: 0
Maximum DB seq length: 2000000000

Post-processing: Minimum Match 0%
Maximum Match 100%
Listing first 45 summaries

Database : Issued_Patents_AA:*
1: /cgn2_6/ptodata/1/iaa/5_COMB.pep:*
2: /cgn2_6/ptodata/1/iaa/6_COMB.pep:*
3: /cgn2_6/ptodata/1/iaa/H_COMB.pep:*
4: /cgn2_6/ptodata/1/iaa/PCTUS_COMB.pep:*
5: /cgn2_6/ptodata/1/iaa/RE_COMB.pep:*
6: /cgn2_6/ptodata/1/iaa/backfiles1.pep:*

Pred. No. is the number of results predicted by chance to have a score greater than or equal to the score of the result being printed, and is derived by analysis of the total score distribution.

SUMMARIES

Result No.	Score	% Query Match	Length	DB	ID	Description
1	551	100.0	639	2	US-10-164-595-44	Sequence 44, Appl
2	202	36.7	141	2	US-09-270-767-41799	Sequence 41799, A
3	202	36.7	141	2	US-09-270-767-57041	Sequence 57041, A
4	73.5	13.3	1495	2	US-08-522-726B-1	Sequence 1, Appli
5	73.5	13.3	1495	2	US-09-337-384-1	Sequence 1, Appli
6	72.5	13.2	642	2	US-09-583-110-3809	Sequence 3809, Ap
7	72.5	13.2	651	2	US-09-107-433-2845	Sequence 2845, Ap
8	69	12.5	223	2	US-08-914-375C-25	Sequence 25, Appl
9	68	12.3	2504	2	US-09-328-352-5821	Sequence 5821, Ap
10	67.5	12.3	150	2	US-09-732-210-1368	Sequence 1368, Ap
11	67.5	12.3	223	2	US-08-914-375C-24	Sequence 24, Appl
12	67.5	12.3	223	2	US-08-914-375C-26	Sequence 26, Appl
13	67.5	12.3	756	2	US-09-248-796A-14799	Sequence 14799, A
14	67	12.2	481	2	US-09-605-703B-2774	Sequence 2774, Ap
15	66.5	12.1	356	2	US-09-248-796A-22834	Sequence 22834, A
16	66	12.0	441	2	US-09-134-000C-6563	Sequence 6563, Ap
17	65.5	11.9	98	2	US-09-328-352-6456	Sequence 6456, Ap
18	65.5	11.9	113	2	US-09-513-999C-8138	Sequence 8138, Ap
19	65.5	11.9	184	2	US-09-248-796A-20343	Sequence 20343, A
20	65.5	11.9	258	2	US-09-270-767-32398	Sequence 32398, A
21	65.5	11.9	357	2	US-09-949-016-8624	Sequence 8624, Ap
22	65.5	11.9	377	2	US-09-949-016-8302	Sequence 8302, Ap
23	65.5	11.9	1113	2	US-09-629-616-3	Sequence 3, Appli
24	65.5	11.9	1113	2	US-10-284-334-3	Sequence 3, Appli
25	65	11.8	1036	2	US-10-104-047-2812	Sequence 2812, Ap
26	65	11.8	1220	2	US-09-540-236-3011	Sequence 3011, Ap
27	65	11.8	2482	2	US-09-252-991A-16967	Sequence 16967, A
28	64.5	11.7	589	2	US-09-809-665A-18	Sequence 18, Appl
29	64.5	11.7	1643	2	US-09-809-665A-103	Sequence 103, App
30	63.5	11.5	168	2	US-09-107-532A-3748	Sequence 3748, Ap
31	63.5	11.5	278	2	US-09-902-540-11503	Sequence 11503, A
32	63.5	11.5	324	2	US-09-328-352-7458	Sequence 7458, Ap
33	63.5	11.5	1088	1	US-08-742-026-2	Sequence 2, Appli
34	63.5	11.5	1088	1	US-08-742-026-23	Sequence 23, Appl
35	63	11.4	318	2	US-09-206-551-54	Sequence 54, Appl
36	63	11.4	492	2	US-08-724-466B-2	Sequence 2, Appli
37	63	11.4	492	2	US-08-882-164D-2	Sequence 2, Appli
38	63	11.4	492	2	US-09-668-482-2	Sequence 2, Appli
39	63	11.4	518	2	US-09-206-551-45	Sequence 45, Appl
40	62.5	11.3	178	1	US-08-689-916A-2	Sequence 2, Appli
41	62.5	11.3	206	2	US-09-248-796A-18593	Sequence 18593, A
42	62.5	11.3	213	2	US-09-252-991A-21520	Sequence 21520, A
43	62.5	11.3	343	2	US-10-104-047-3068	Sequence 3068, Ap
44	62.5	11.3	502	2	US-09-248-796A-15481	Sequence 15481, A
45	62.5	11.3	910	2	US-09-623-326-7	Sequence 7, Appli

ALIGNMENTS

RESULT 1
US-10-164-595-44
; Sequence 44, Application US/10164595
; Patent No. 6657054
; GENERAL INFORMATION:
; APPLICANT: OriGene Technologies, Inc
; TITLE OF INVENTION: Regulated Angiogenesis Genes and Polypeptides
; FILE REFERENCE: 1U 103 R1
; CURRENT APPLICATION NUMBER: US/10/164,595
; CURRENT FILING DATE: 2002-06-10
; NUMBER OF SEQ ID NOS: 80
; SOFTWARE: PatentIn version 3.1
; SEQ ID NO 44
; LENGTH: 639
; TYPE: PRT
; ORGANISM: Homo sapiens
US-10-164-595-44

Query Match 100.0%; Score 551; DB 2; Length 639;
Best Local Similarity 100.0%; Pred. No. 2.3e-60;
Matches 108; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

```
Qy          1 KLAFRACRLKKKAQYEANKVKLWGLNTEYDNLLFVINSIKQEIVNRVQNPRDERGPNMGQ 60
              ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db        532 KLAFRACRLKKKAQYEANKVKLWGLNTEYDNLLFVINSIKQEIVNRVQNPRDERGPNMGQ 591

Qy         61 KLEILIKDTLGLPVAGQTSEFVNQVLEKTAEGNPTGGLVGLRIPTSKV 108
              ||||||||||||||||||||||||||||||||||||||||||||||||||
Db        592 KLEILIKDTLGLPVAGQTSEFVNQVLEKTAEGNPTGGLVGLRIPTSKV 639
```

RESULT 2
US-09-270-767-41799
; Sequence 41799, Application US/09270767
; Patent No. 6703491
; GENERAL INFORMATION:
; APPLICANT: Homburger et al.
; TITLE OF INVENTION: Nucleic acids and proteins of Drosophila melanogaster
; FILE REFERENCE: File Reference: 7326-094
; CURRENT APPLICATION NUMBER: US/09/270,767
; CURRENT FILING DATE: 1999-03-17
; NUMBER OF SEQ ID NOS: 62517
; SOFTWARE: PatentIn Ver. 2.0
; SEQ ID NO 41799
; LENGTH: 141
; TYPE: PRT
; ORGANISM: Drosophila melanogaster
US-09-270-767-41799

Query Match 36.7%; Score 202; DB 2; Length 141;

```
  Best Local Similarity   43.6%;  Pred. No. 2.2e-17;
  Matches   48;  Conservative   17;  Mismatches   31;  Indels   14;  Gaps    2;

Qy          1 KLAFRACRLKKKAQYEANKVKLWGLNTEYDNLLFVINSIKQEIVNRVQNPRDERGPNMGQ 60
              ||| |||||||||||:||||:||:||  |:  |:  |  :|| :| :       |  |:|:
Db         35 KLASRACRLKKKAQHEANKIKLFGLEIEHKRLMNGIAELKQALVVK------HRTKNLGE 88

Qy         61 KLE--------ILIKDTLGLPVAGQTSEFVNQVLEKTAEGNPTGGLVGLR 102
                |       |    : |: :|| :::|||:|||    | | |||  ||
Db         89 STEEVDQQIARIYATASSGIRIAGGSTDFVNKVLENMRGGMPNGGLEELR 138


RESULT 3
US-09-270-767-57041
; Sequence 57041, Application US/09270767
; Patent No. 6703491
; GENERAL INFORMATION:
;  APPLICANT: Homburger et al.
;  TITLE OF INVENTION: Nucleic acids and proteins of Drosophila melanogaster
;  FILE REFERENCE: File Reference: 7326-094
;  CURRENT APPLICATION NUMBER: US/09/270,767
;  CURRENT FILING DATE: 1999-03-17
;  NUMBER OF SEQ ID NOS: 62517
;  SOFTWARE: PatentIn Ver. 2.0
; SEQ ID NO 57041
;   LENGTH: 141
;   TYPE: PRT
;   ORGANISM: Drosophila melanogaster
US-09-270-767-57041

  Query Match             36.7%;  Score 202;  DB 2;  Length 141;
  Best Local Similarity   43.6%;  Pred. No. 2.2e-17;
  Matches   48;  Conservative   17;  Mismatches   31;  Indels   14;  Gaps    2;

Qy          1 KLAFRACRLKKKAQYEANKVKLWGLNTEYDNLLFVINSIKQEIVNRVQNPRDERGPNMGQ 60
              ||| |||||||||||:||||:||:||  |:  |:  |  :|| :| :       |  |:|:
Db         35 KLASRACRLKKKAQHEANKIKLFGLEIEHKRLMNGIAELKQALVVK------HRTKNLGE 88

Qy         61 KLE--------ILIKDTLGLPVAGQTSEFVNQVLEKTAEGNPTGGLVGLR 102
                |       |    : |: :|| :::|||:|||    | | |||  ||
Db         89 STEEVDQQIARIYATASSGIRIAGGSTDFVNKVLENMRGGMPNGGLEELR 138


RESULT 4
US-08-522-726B-1
; Sequence 1, Application US/08522726B
; Patent No. 6489441
;  GENERAL INFORMATION:
;    APPLICANT:  Evans, Ronald M.
;    APPLICANT:  Chen, J. Don
;    TITLE OF INVENTION:  TRANSCRIPTIONAL CO-REPRESSOR THAT
;    TITLE OF INVENTION:  INTERACTS WITH NUCLEAR HORMONE RECEPTORS AND USES
;    TITLE OF INVENTION:  THEREFOR
;    NUMBER OF SEQUENCES:  3
;    CORRESPONDENCE ADDRESS:
;      ADDRESSEE:  Pretty, Schroeder, Brueggemann & Clark
;      STREET:  444 South Flower Street, Suite 2000
;      CITY:  Los Angeles
;      STATE:  CA
;      COUNTRY:  USA
;      ZIP:  90071
;    COMPUTER READABLE FORM:
;      MEDIUM TYPE:  Floppy disk
;      COMPUTER:  IBM PC compatible
;      OPERATING SYSTEM:  PC-DOS/MS-DOS
;      SOFTWARE:  PatentIn Release #1.0, Version #1.25
;    CURRENT APPLICATION DATA:
;      APPLICATION NUMBER:  US/08/522,726B
;      FILING DATE:  01-SEP-1995
;      CLASSIFICATION:  435
;    ATTORNEY/AGENT INFORMATION:
;      NAME:  Reiter, Stephen E.
;      REGISTRATION NUMBER:  31,192
;      REFERENCE/DOCKET NUMBER:  P41 90042
;    TELECOMMUNICATION INFORMATION:
;      TELEPHONE:  619-546-1995
;      TELEFAX:  619-546-9392
;  INFORMATION FOR SEQ ID NO:  1:
;    SEQUENCE CHARACTERISTICS:
;      LENGTH:  1495 amino acids
;      TYPE:  amino acid
;      STRANDEDNESS:  single
;      TOPOLOGY:  both
;    MOLECULE TYPE:  protein
US-08-522-726B-1

  Query Match             13.3%;  Score 73.5;  DB 2;  Length 1495;
  Best Local Similarity   30.6%;  Pred. No. 9.1;
  Matches   22;  Conservative   14;  Mismatches   17;  Indels   19;  Gaps    3;

Qy         37 NSIKQEIVNRVQNPRDERGPNMGQKLEILIKDTLGLPVAGQTSEFVNQVLEKTAEGNPTG 96
              : :||| :    :|| : ||           ::||:| | : |     ||  || |:  |
Db        143 SGVKQEQL----SPRGQAGP----------PESLGVPTAQEAS-----VLRGTALGSVPG 183

Qy         97 GLVGLRIPTSKV 108
              | :   ||:::|
Db        184 GSITKGIPSTRV 195


RESULT 5
US-09-337-384-1
; Sequence 1, Application US/09337384
; Patent No. 6551773
; GENERAL INFORMATION:
;  APPLICANT:  THE SALK INSTITUTE FOR BIOLOGICAL STUDIES
;  APPLICANT:  EVANS, RONALD
;  APPLICANT:  CHEN, J.
;  TITLE OF INVENTION:  TRANSCRIPTIONAL CO-REPRESSOR THAT INTERACTS WITH NUCLEAR HORMONI
;  TITLE OF INVENTION:  RECEPTORS
;  FILE REFERENCE:  SALK1510-2
;  CURRENT APPLICATION NUMBER: US/09/337,384
;  CURRENT FILING DATE:  1999-06-21
;  PRIOR APPLICATION NUMBER:  08/522,726
;  PRIOR FILING DATE:  1995-09-01
;  NUMBER OF SEQ ID NOS:  3
;  SOFTWARE:  PatentIn version 3.0
; SEQ ID NO 1
;   LENGTH: 1495
;   TYPE: PRT
;   ORGANISM: Homo sapiens
US-09-337-384-1

  Query Match             13.3%;  Score 73.5;  DB 2;  Length 1495;
  Best Local Similarity   30.6%;  Pred. No. 9.1;
  Matches   22;  Conservative   14;  Mismatches   17;  Indels   19;  Gaps    3;

Qy         37 NSIKQEIVNRVQNPRDERGPNMGQKLEILIKDTLGLPVAGQTSEFVNQVLEKTAEGNPTG 96
              : :||| :    :|| : ||           ::||:| | : |     ||  || |:  |
Db        143 SGVKQEQL----SPRGQAGP----------PESLGVPTAQEAS-----VLRGTALGSVPG 183

Qy         97 GLVGLRIPTSKV 108
              | :   ||:::|
Db        184 GSITKGIPSTRV 195


RESULT 6
US-09-583-110-3809
; Sequence 3809, Application US/09583110
; Patent No. 6699703
; GENERAL INFORMATION:
;  APPLICANT:  Lynn Doucette-Stamm et al.
;  TITLE OF INVENTION:  Nucleic Acid and Amino Acid Sequences Relating to Streptococcus
;  TITLE OF INVENTION:  Pneumoniae for Diagnostics and Therapeutics
;  FILE REFERENCE:  PATH00-07A
;  CURRENT APPLICATION NUMBER: US/09/583,110
;  CURRENT FILING DATE:  2000-05-26
```

```
;  PRIOR APPLICATION NUMBER:  US 09/107,433
;  PRIOR FILING DATE:  1998-06-30
;  PRIOR APPLICATION NUMBER:  US 60/085,131
;  PRIOR FILING DATE:  1998-05-12
;  PRIOR APPLICATION NUMBER:  US 60/051,553
;  PRIOR FILING DATE:  1997-07-02
;  NUMBER OF SEQ ID NOS:  5322
; SEQ ID NO 3809
;   LENGTH: 642
;   TYPE: PRT
;   ORGANISM: Streptococcus pneumoniae
US-09-583-110-3809

  Query Match             13.2%;  Score 72.5;  DB 2;  Length 642;
  Best Local Similarity   31.0%;  Pred. No. 3.7;
  Matches   26;  Conservative   12;  Mismatches   15;  Indels   31;  Gaps    5;

Qy         22 LWGLNTEYDNLLFVINSIKQEIVNRVQNPRDERGPNM------GQKLEILI--------- 66
              || ::| || :          ||::      | |  ::      |:| |:|:
Db        511 LWEIDTSYDGI---------EIID-----ADNRDQSVLSFIRKGKKGEMLVCIFNMVPVE 556

Qy         67 -KD-TLGLPVAGQTSEFVNQVLEK 88
               || |:||||||    |  |  ||:
Db        557 RKDFTIGLPVAGIYEEVWNTELEE 580


RESULT 7
US-09-107-433-2845
; Sequence 2845, Application US/09107433
; Patent No. 6800744
;   GENERAL INFORMATION:
;        APPLICANT: Lynn A Doucette-Stamm  and David Bush
;        TITLE OF INVENTION: NUCLEIC ACID AND AMINO ACID
;                            SEQUENCES RELATING TO STREPTOCOCCUS PNEUMONIAE   FOR DIAGNO
;                            THERAPEUTICS
;        NUMBER OF SEQUENCES: 5206
;        CORRESPONDENCE ADDRESS:
;             ADDRESSEE: GENOME THERAPEUTICS CORPORATION
;             STREET: 100 Beaver Street
;             CITY: Waltham
;             STATE: Massachusetts
;             COUNTRY: USA
;             ZIP: 02354
;        COMPUTER READABLE FORM:
;             MEDIUM TYPE: CD/ROM ISO9660
;             COMPUTER: <Unknown>
;             OPERATING SYSTEM: <Unknown>
;             SOFTWARE: <Unknown>
;        CURRENT APPLICATION DATA:
;             APPLICATION NUMBER: US/09/107,433
;             FILING DATE: 30-Jun-1998
;        PRIOR APPLICATION DATA:
;             APPLICATION NUMBER: 60/ 085131
;             FILING DATE: May 12, 1998
;             APPLICATION NUMBER: 60/051553
;             FILING DATE: July 2, 1997
;        ATTORNEY/AGENT INFORMATION:
;             NAME: Ariniello, Pamela Deneke
;             REGISTRATION NUMBER: 40,489
;             REFERENCE/DOCKET NUMBER: GTC-011
;        TELECOMMUNICATION INFORMATION:
;             TELEPHONE: (781)893-5007
;             TELEFAX: (781)893-8277
;   INFORMATION FOR SEQ ID NO: 2845:
;        SEQUENCE CHARACTERISTICS:
;             LENGTH: 651 amino acids
;             TYPE: amino acid
;             TOPOLOGY: linear
;        MOLECULE TYPE: protein
;        HYPOTHETICAL: YES
;        ORIGINAL SOURCE:
;             ORGANISM: Streptococcus pneumoniae
;        FEATURE:
;             NAME/KEY:  misc_feature
;             LOCATION: (B) LOCATION 1...651
;        SEQUENCE DESCRIPTION: SEQ ID NO: 2845:
US-09-107-433-2845

  Query Match             13.2%;  Score 72.5;  DB 2;  Length 651;
  Best Local Similarity   31.0%;  Pred. No. 3.7;
  Matches   26;  Conservative   12;  Mismatches   15;  Indels   31;  Gaps    5;

Qy         22 LWGLNTEYDNLLFVINSIKQEIVNRVQNPRDERGPNM------GQKLEILI--------- 66
              || ::| || :          ||::      | |  ::      |:| |:|:
Db        520 LWEIDTSYDGI---------EIID-----ADNRDQSVLSFIRKGKKGEMLVCIFNMVPVE 565

Qy         67 -KD-TLGLPVAGQTSEFVNQVLEK 88
               || |:||||||    |  |  ||:
Db        566 RKDFTIGLPVAGIYEEVWNTELEE 589


RESULT 8
US-08-914-375C-25
; Sequence 25, Application US/08914375C
; Patent No. 6377893
;   GENERAL INFORMATION:
;        APPLICANT: Steven A. Benner
;                   Applications of Protein Structure Predictions
;        NUMBER OF SEQUENCES: 74
;        CORRESPONDENCE ADDRESS:
;             ADDRESSEE: Steven A. Benner
;             STREET: 1501 NW 68th Terrace
;             CITY: Gainesville
;             STATE: FL
;             COUNTRY: United States
;             ZIP: 32605-4147
;        COMPUTER READABLE FORM:
;             MEDIUM TYPE: 3.5 inch diskette
;             COMPUTER: Apple MacIntosh
;             OPERATING SYSTEM: MacIntosh 7.0
;             SOFTWARE: Microsoft Word
;        CURRENT APPLICATION DATA:
;             APPLICATION NUMBER: US/08/914,375C
;             FILING DATE: 19-Aug-1997
;             CLASSIFICATION: 702/20
;        TELECOMMUNICATION INFORMATION:
;             TELEPHONE: 352 392 7773
;             TELEFAX: 352 331 0462
;   INFORMATION FOR SEQ ID NO: 25:
;        SEQUENCE CHARACTERISTICS:
;             LENGTH: 223
;             TYPE: amino acid
;             TOPOLOGY: linear
;        MOLECULE TYPE: amino acid
;        ORIGINAL SOURCE:
;             ORGANISM: Homo sapiens
;        FEATURE:
;             OTHER INFORMATION: ENPL_HUMAN ENDOPLASMIN PRECURSOR (94 KD GLUCOSE- REGULAT
;             PROTEIN) (GRP94)
;        SEQUENCE DESCRIPTION: SEQ ID NO: 25:
US-08-914-375C-25

  Query Match             12.5%;  Score 69;  DB 2;  Length 223;
  Best Local Similarity   27.3%;  Pred. No. 2.3;
  Matches   24;  Conservative   19;  Mismatches   33;  Indels   12;  Gaps    4;

Qy         18 NKVKLWGLNTEYDNLLFVINSIKQEIVNRVQNPRDER-----GPNMGQKLEILIKDTLGL 72
              :|::|  |  |  | |     | :|:  :::  :::         :|   | |:|: ||
Db         43 DKIRLISLTDE--NAL----SGNEELTVKIKCDKEKNLLHVTDTGVGMTREELVKN-LGT 95

Qy         73 PVAGQTSEFVNQVLEKTAEGNPTGGLVG 100
                  ||||:|:: |   :|  |  |:|
Db         96 IAKSGTSEFLNKMTEAQEDGQSTSELIG 123


RESULT 9
```

```
US-09-328-352-5821
; Sequence 5821, Application US/09328352
; Patent No. 6562958
; GENERAL INFORMATION:
;  APPLICANT: Gary L. Breton et al.
;  TITLE OF INVENTION: NUCLEIC ACID AND AMINO ACID SEQUENCES RELATING TO ACINETOBACTER
;  TITLE OF INVENTION: BAUMANNII FOR DIAGNOSTICS AND THERAPEUTICS
;  FILE REFERENCE: GTC99-03PA
;  CURRENT APPLICATION NUMBER: US/09/328,352
;  CURRENT FILING DATE: 1999-06-04
;  NUMBER OF SEQ ID NOS: 8252
; SEQ ID NO 5821
;   LENGTH: 2504
;   TYPE: PRT
;   ORGANISM: Acinetobacter baumannii
US-09-328-352-5821

  Query Match             12.3%;  Score 68;  DB 2;  Length 2504;
  Best Local Similarity   26.6%;  Pred. No. 93;
  Matches   25;  Conservative   10;  Mismatches   31;  Indels   28;  Gaps    3;

Qy         25 LNTEYDNLLFVINSIKQEIVNRVQNPRDE--------------RGPN------------M 58
              |||  | |   ||  | |: | : | : |              :| |            :
Db       1671 LNTTKDQLTTQINDTKTELNNTIGNTKTELNSKIDSTKTELENKGLNFAGNSGNDVHRKL 1730

Qy         59 GQKLEILIKDTLGLPVAGQTSEFVNQVLEKTAEG 92
              |:|| |:       |||  : |  | :   | :|
Db       1731 GEKLNIIGGAAASTPVAKTSGE--NVITRTTQDG 1762


RESULT 10
US-09-732-210-1368
; Sequence 1368, Application US/09732210
; Patent No. 6573361
; GENERAL INFORMATION:
;  APPLICANT: Bunkers, Greg J.
;  APPLICANT:  Liang, Jihong
;  APPLICANT:  Mittanck, Cindy A.
;  APPLICANT:  Seale, Jeffrey W.
;  APPLICANT:  Wu, Yonnie S.
;  TITLE OF INVENTION: Anti-fungal Proteins and Methods for Their Use
;  FILE REFERENCE: 38-21(15036)B
;  CURRENT APPLICATION NUMBER: US/09/732,210
;  CURRENT FILING DATE:  2000-12-07
;  PRIOR APPLICATION NUMBER: US 60/169,513
;  PRIOR FILING DATE: 1999-12-07
;  PRIOR APPLICATION NUMBER: US 60/169,340
;  PRIOR FILING DATE: 1999-12-07
;  NUMBER OF SEQ ID NOS: 1753
; SEQ ID NO 1368
;   LENGTH: 150
;   TYPE: PRT
;   ORGANISM: Methanococcus jannaschii
US-09-732-210-1368

  Query Match             12.3%;  Score 67.5;  DB 2;  Length 150;
  Best Local Similarity   29.4%;  Pred. No. 2;
  Matches   30;  Conservative   11;  Mismatches   32;  Indels   29;  Gaps    5;

Qy          9 LKKKAQYEANKVKL----WGLNTEYDNLL----FVINS----IKQEIVNRVQNPRDERGP 56
              :||  :   |:  |     |  |   | :      || |    ||||  :||::  |  ||
Db         61 VKKIEEVLADPAKFGIPSWMFNRRKDYVTGEDKHVIESDLMIIKQEDINRLKRIRCYRG- 119

Qy         57 NMGQKLEILIKDTLGLPVAGQTSEFVNQVLEKTAEGNPTGGL 98
                        |:  ||||  || :       : |    || |:
Db        120 ---------IRHELGLPCRGQRT-------KSTFRRGPTVGV 145


RESULT 11
US-08-914-375C-24
; Sequence 24, Application US/08914375C
; Patent No. 6377893
;   GENERAL INFORMATION:
;       APPLICANT: Steven A. Benner
;                  Applications of Protein Structure Predictions
;       NUMBER OF SEQUENCES: 74
;       CORRESPONDENCE ADDRESS:
;            ADDRESSEE: Steven A. Benner
;            STREET: 1501 NW 68th Terrace
;            CITY: Gainesville
;            STATE: FL
;            COUNTRY: United States
;            ZIP: 32605-4147
;       COMPUTER READABLE FORM:
;            MEDIUM TYPE: 3.5 inch diskette
;            COMPUTER: Apple MacIntosh
;            OPERATING SYSTEM: MacIntosh 7.0
;            SOFTWARE: Microsoft Word
;       CURRENT APPLICATION DATA:
;            APPLICATION NUMBER: US/08/914,375C
;            FILING DATE: 19-Aug-1997
;            CLASSIFICATION: 702/20
;       TELECOMMUNICATION INFORMATION:
;            TELEPHONE: 352 392 7773
;            TELEFAX: 352 331 0462
;   INFORMATION FOR SEQ ID NO: 24:
;       SEQUENCE CHARACTERISTICS:
;            LENGTH: 223
;            TYPE: amino acid
;            TOPOLOGY: linear
;       MOLECULE TYPE: amino acid
;       ORIGINAL SOURCE:
;            ORGANISM: Canis familiaris
;       FEATURE:
;            OTHER INFORMATION: ENPL_CANFA ENDOPLASMIN PRECURSOR (94 KD GLUCOSE- REGULA
;            PROTEIN) (GRP94)
;       SEQUENCE DESCRIPTION: SEQ ID NO: 24:
US-08-914-375C-24

  Query Match             12.3%;  Score 67.5;  DB 2;  Length 223;
  Best Local Similarity   20.2%;  Pred. No. 3.5;
  Matches   26;  Conservative   23;  Mismatches   41;  Indels   39;  Gaps    4;

Qy          1 KLAFRACRLKKKAQYEANKVKLWGLNTEYDNLLFVINSI---------------------- 39
              | ||:|          | |::    :|: | |    :  :
Db          5 KFAFQA---------EVNRMMKLIINSLYKNKEIFLRELISNASDALDKIRLISLTDENA 55

Qy         40 ---KQEIVNRVQNPRDER-----GPNMGQKLEILIKDTLGLPVAGQTSEFVNQVLEKTAE 91
                 :|:  :::  :::          :|   | |:|: ||       ||||:|:: |   :
Db         56 LAGNEELTVKIKCDKEKNLLHVTDTGVGMTREELVKN-LGTIAKSGTSEFLNKMTEAQED 114

Qy         92 GNPTGGLVG 100
              |  |  |:|
Db        115 GQSTSELIG 123


RESULT 12
US-08-914-375C-26
; Sequence 26, Application US/08914375C
; Patent No. 6377893
;   GENERAL INFORMATION:
;       APPLICANT: Steven A. Benner
;                  Applications of Protein Structure Predictions
;       NUMBER OF SEQUENCES: 74
;       CORRESPONDENCE ADDRESS:
;            ADDRESSEE: Steven A. Benner
;            STREET: 1501 NW 68th Terrace
;            CITY: Gainesville
;            STATE: FL
;            COUNTRY: United States
;            ZIP: 32605-4147
;       COMPUTER READABLE FORM:
;            MEDIUM TYPE: 3.5 inch diskette
;            COMPUTER: Apple MacIntosh
;            OPERATING SYSTEM: MacIntosh 7.0
;            SOFTWARE: Microsoft Word
;
```

```
;       CURRENT APPLICATION DATA:
;            APPLICATION NUMBER: US/08/914,375C
;            FILING DATE: 19-Aug-1997
;            CLASSIFICATION: 702/20
;       TELECOMMUNICATION INFORMATION:
;            TELEPHONE: 352 392 7773
;            TELEFAX: 352 331 0462
;   INFORMATION FOR SEQ ID NO: 26:
;       SEQUENCE CHARACTERISTICS:
;            LENGTH: 223
;            TYPE: amino acid
;            TOPOLOGY: linear
;       MOLECULE TYPE: amino acid
;       ORIGINAL SOURCE:
;            ORGANISM: Mus musculus
;       FEATURE:
;            OTHER INFORMATION: ENPL_MOUSE ENDOPLASMIN PRECURSOR (94 KD GLUCOSE- P11427
;            REGULATED PROTEIN) (GRP94)
;       SEQUENCE DESCRIPTION: SEQ ID NO: 26:
US-08-914-375C-26

  Query Match             12.3%;  Score 67.5;  DB 2;  Length 223;
  Best Local Similarity   20.2%;  Pred. No. 3.5;
  Matches   26;  Conservative   23;  Mismatches   41;  Indels   39;  Gaps    4;

Qy          1 KLAFRACRLKKKAQYEANKVKLWGLNTEYDNLLFVINSI---------------------- 39
              | ||:|          | |::    :|: | |     :  :
Db          5 KFAFQA---------EVNRMMKLIINSLYKNKEIFLRELISNASDALDKIRLISLTDENA 55

Qy         40 ---KQEIVNRVQNPRDER-----GPNMGQKLEILIKDTLGLPVAGQTSEFVNQVLEKTAE 91
                 :|:  :::  :::         :|   | |:|: ||      ||||:|:: |   :
Db         56 LAGNEELTVKIKCDKEKNLLHVTDTGVGMTREELVKN-LGTIAKSGTSEFLNKMTEAQED 114

Qy         92 GNPTGGLVG 100
              |  |  |:|
Db        115 GQSTSELIG 123


RESULT 13
US-09-248-796A-14799
; Sequence 14799, Application US/09248796A
; Patent No. 6747137
; GENERAL INFORMATION:
;  APPLICANT: Keith Weinstock et al
;  TITLE OF INVENTION: NUCLEIC ACID AND AMINO ACID SEQUENCES RELATING TO CANDIDA ALBICAN
;  TITLE OF INVENTION:  FOR DIAGNOSTICS AND THERAPEUTICS
;  FILE REFERENCE: 107196.132
;  CURRENT APPLICATION NUMBER: US/09/248,796A
;  CURRENT FILING DATE:  1999-02-12
;  PRIOR APPLICATION NUMBER: US 60/074,725
;  PRIOR FILING DATE: 1998-02-13
;  PRIOR APPLICATION NUMBER: US 60/096,409
;  PRIOR FILING DATE: 1998-08-13
;  NUMBER OF SEQ ID NOS: 28208
; SEQ ID NO 14799
;   LENGTH: 756
;   TYPE: PRT
;   ORGANISM: Candida albicans
US-09-248-796A-14799

  Query Match             12.3%;  Score 67.5;  DB 2;  Length 756;
  Best Local Similarity   30.0%;  Pred. No. 20;
  Matches   27;  Conservative   13;  Mismatches   23;  Indels   27;  Gaps    6;

Qy         25 LNTE-YD--NLLFVINSI----KQEIVNRVQNPRDER------GPNMGQKLEILIKDTLG 71
              :||| ||  ||:   |||    |:|| |::   |  |      | ::  |:::
Db        427 INTELYDKVNLISNPNSIYDQRKREIANKINEERKSRILTSSNGNDLPTKIKV------- 479

Qy         72 LPVAGQTSEFVNQVLEKTAE-GNPTGGLVG 100
                    : ||::  | || | | |   |
Db        480 ------NKDLVNKLQTKFAENGTPDGNANG 503


RESULT 14
US-09-605-703B-2774
; Sequence 2774, Application US/09605703B
; Patent No. 6962989
; GENERAL INFORMATION:
;  APPLICANT: Pompejus, Markus
;  APPLICANT:  Kroger, Burkhard
;  APPLICANT:  Schroder, Hartwig
;  APPLICANT:  Zelder, Oskar
;  APPLICANT:  Haberhauer, Gregor
;  TITLE OF INVENTION: CORYNEBACTERIUM GLUTAMICUM GENES ENCODING NOVEL
;  TITLE OF INVENTION:  PROTEINS
;  FILE REFERENCE: BGI-129CP
;  CURRENT APPLICATION NUMBER: US/09/605,703B
;  CURRENT FILING DATE:  2000-06-27
;  PRIOR APPLICATION NUMBER: 60/142,764
;  PRIOR FILING DATE: 1999-07-08
;  PRIOR APPLICATION NUMBER: 60/152,318
;  PRIOR FILING DATE: 1999-09-03
;  NUMBER OF SEQ ID NOS: 2934
; SEQ ID NO 2774
;   LENGTH: 481
;   TYPE: PRT
;   ORGANISM: Corynebacterium glutamicum
US-09-605-703B-2774

  Query Match             12.2%;  Score 67;  DB 2;  Length 481;
  Best Local Similarity   22.8%;  Pred. No. 12;
  Matches   21;  Conservative   21;  Mismatches   24;  Indels   26;  Gaps    4;

Qy          8 RLKKKAQYEANKVKLWGLN---TEYDNLLFVINSIKQEIVNRVQNPRDERGPNMGQKLEI 64
              |||  : ::   |  ||  :    ::||: : : | :::  :     | |
Db         66 RLKALSIFDKKPVPTWGADLSGIDFDNIKYFVRSTEKQAQSWEDLPED------------ 113

Qy         65 LIKDT---LGLP-------VAGQTSEFVNQVL 86
                ||:|   ||:|       |||  ::: ::|:
Db        114 -IKNTYDKLGIPEAEKQRLVAGVAAQYESEVV 144


RESULT 15
US-09-248-796A-22834
; Sequence 22834, Application US/09248796A
; Patent No. 6747137
; GENERAL INFORMATION:
;  APPLICANT: Keith Weinstock et al
;  TITLE OF INVENTION: NUCLEIC ACID AND AMINO ACID SEQUENCES RELATING TO CANDIDA ALBICANS
;  TITLE OF INVENTION:  FOR DIAGNOSTICS AND THERAPEUTICS
;  FILE REFERENCE: 107196.132
;  CURRENT APPLICATION NUMBER: US/09/248,796A
;  CURRENT FILING DATE:  1999-02-12
;  PRIOR APPLICATION NUMBER: US 60/074,725
;  PRIOR FILING DATE: 1998-02-13
;  PRIOR APPLICATION NUMBER: US 60/096,409
;  PRIOR FILING DATE: 1998-08-13
;  NUMBER OF SEQ ID NOS: 28208
; SEQ ID NO 22834
;   LENGTH: 356
;   TYPE: PRT
;   ORGANISM: Candida albicans
US-09-248-796A-22834

  Query Match             12.1%;  Score 66.5;  DB 2;  Length 356;
  Best Local Similarity   21.1%;  Pred. No. 9.1;
  Matches   19;  Conservative   19;  Mismatches   31;  Indels   21;  Gaps    3;

Qy         13 AQYEANKVKLW----GLNTEYDNLLF-------------VINSIKQEIVNRVQNPRDER 54
              | |  :|  :|    |:|| |:  |               : | ::  | | ::  ::|:
Db        207 ANYYLSKEYIWILYSGVNTYYERFGFKSFPMDFYEVPNTLLTNDVESAIENLIKESKEEK 266

Qy         55 GPNMGQKLEILIKDTLGLPVAGQTSEFVNQ 84
                 |: : :|: |    |   :  :|: |
Db        267 QLPSGKSIRLLLGDN---PADQEIIQFILQ 293
```

Search completed: February 28, 2006, 08:56:08
Job time : 9.54323 secs

GenCore version 5.1.7

OM protein - protein search, using sw model

Run on: February 28, 2006, 08:54:36 ; Search time 31.4887 Seconds
(without alignments)
1433.071 Million cell updates/sec

Title: US-10-717-665A-44_COPY_532_639
Perfect score: 551
Sequence: 1 KLAFRACRLKKKAQYEANKV.....TAEGNPTGGLVGLRIPTSKV 108

Scoring table: BLOSUM62
Gapop 10.0 , Gapext 0.5

Searched: 1867569 seqs, 417829326 residues

Total number of hits satisfying chosen parameters: 1867569

Minimum DB seq length: 0
Maximum DB seq length: 2000000000

Post-processing: Minimum Match 0%
Maximum Match 100%
Listing first 45 summaries

Database : Published_Applications_AA_Main:*
1: /cgn2_6/ptodata/1/pubpaa/US07_PUBCOMB.pep:*
2: /cgn2_6/ptodata/1/pubpaa/US08_PUBCOMB.pep:*
3: /cgn2_6/ptodata/1/pubpaa/US09_PUBCOMB.pep:*
4: /cgn2_6/ptodata/1/pubpaa/US10A_PUBCOMB.pep:*
5: /cgn2_6/ptodata/1/pubpaa/US10B_PUBCOMB.pep:*
6: /cgn2_6/ptodata/1/pubpaa/US11_PUBCOMB.pep:*

Pred. No. is the number of results predicted by chance to have a score greater than or equal to the score of the result being printed, and is derived by analysis of the total score distribution.

SUMMARIES

Result No.	Score	% Query Match	Length	DB	ID	Description
1	551	100.0	639	5	US-10-717-665-44	Sequence 44, Appl
2	202	36.7	814	6	US-11-097-143-31914	Sequence 31914, A
3	200.5	36.4	160	5	US-10-450-763-59634	Sequence 59634, A
4	78.5	14.2	281	4	US-10-282-122A-47201	Sequence 47201, A
5	75.5	13.7	345	5	US-10-472-928-2234	Sequence 2234, Ap
6	75.5	13.7	1175	4	US-10-369-493-22667	Sequence 22667, A
7	74.5	13.5	1124	4	US-10-369-493-19968	Sequence 19968, A
8	73.5	13.3	625	4	US-10-282-122A-71587	Sequence 71587, A
9	73.5	13.3	1104	4	US-10-369-493-19087	Sequence 19087, A
10	73.5	13.3	1495	4	US-10-351-750-1	Sequence 1, Appli
11	73.5	13.3	1495	5	US-10-705-165-26	Sequence 26, Appl
12	73.5	13.3	2507	3	US-09-819-104A-2	Sequence 2, Appli
13	73.5	13.3	2507	5	US-10-723-860-1740	Sequence 1740, Ap
14	73.5	13.3	2517	4	US-10-087-192-654	Sequence 654, App
15	72.5	13.2	651	5	US-10-617-320-2845	Sequence 2845, Ap
16	72.5	13.2	4322	4	US-10-437-963-104793	Sequence 104793,
17	72	13.1	491	4	US-10-425-115-315525	Sequence 315525,
18	72	13.1	861	4	US-10-369-493-3509	Sequence 3509, Ap
19	71	12.9	505	6	US-11-097-143-23307	Sequence 23307, A
20	71	12.9	825	4	US-10-369-493-8932	Sequence 8932, Ap
21	71	12.9	826	5	US-10-732-923-6933	Sequence 6933, Ap
22	71	12.9	1355	4	US-10-369-493-5251	Sequence 5251, Ap
23	71	12.9	1355	4	US-10-369-493-5252	Sequence 5252, Ap
24	70.5	12.8	300	4	US-10-437-963-176866	Sequence 176866,
25	70	12.7	487	4	US-10-203-927A-8	Sequence 8, Appli
26	70	12.7	597	4	US-10-369-493-3265	Sequence 3265, Ap
27	70	12.7	846	4	US-10-203-927A-12	Sequence 12, Appl
28	69.5	12.6	1105	4	US-10-369-493-2816	Sequence 2816, Ap
29	69	12.5	220	4	US-10-415-868-2	Sequence 2, Appli
30	69	12.5	282	4	US-10-282-122A-74571	Sequence 74571, A
31	69	12.5	316	3	US-09-968-436B-4	Sequence 4, Appli
32	69	12.5	316	4	US-10-260-104B-4	Sequence 4, Appli
33	69	12.5	339	4	US-10-781-014-380	Sequence 380, App
34	69	12.5	349	3	US-09-738-626-5897	Sequence 5897, Ap
35	69	12.5	395	4	US-10-369-493-8103	Sequence 8103, Ap
36	69	12.5	666	4	US-10-264-049-2706	Sequence 2706, Ap
37	69	12.5	719	4	US-10-408-765A-2106	Sequence 2106, Ap
38	69	12.5	803	3	US-09-759-010-7	Sequence 7, Appli
39	69	12.5	803	3	US-09-968-436B-2	Sequence 2, Appli
40	69	12.5	803	4	US-10-233-553-3	Sequence 3, Appli
41	69	12.5	803	4	US-10-260-104B-2	Sequence 2, Appli
42	69	12.5	803	4	US-10-408-765A-514	Sequence 514, App
43	69	12.5	803	5	US-10-733-969A-79	Sequence 79, Appl
44	69	12.5	803	5	US-10-789-378-78	Sequence 78, Appl
45	69	12.5	803	5	US-10-844-711-2	Sequence 2, Appli

ALIGNMENTS

RESULT 1
US-10-717-665-44
; Sequence 44, Application US/10717665
; Publication No. US20050106579A1
; GENERAL INFORMATION:
; APPLICANT: OriGene Technologies, Inc
; TITLE OF INVENTION: Regulated Angiogenesis Genes and Polypeptides
; FILE REFERENCE: 1U 103 R1
; CURRENT APPLICATION NUMBER: US/10/717,665
; CURRENT FILING DATE: 2003-11-21
; PRIOR APPLICATION NUMBER: US/10/164,595
; PRIOR FILING DATE: 2002-06-10
; NUMBER OF SEQ ID NOS: 80
; SOFTWARE: PatentIn version 3.1
; SEQ ID NO 44
; LENGTH: 639
; TYPE: PRT
; ORGANISM: Homo sapiens
US-10-717-665-44

Query Match 100.0%; Score 551; DB 5; Length 639;
Best Local Similarity 100.0%; Pred. No. 3.4e-54;
Matches 108; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

```
Qy          1 KLAFRACRLKKKAQYEANKVKLWGLNTEYDNLLFVINSIKQEIVNRVQNPRDERGPNMGQ 60
              ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db        532 KLAFRACRLKKKAQYEANKVKLWGLNTEYDNLLFVINSIKQEIVNRVQNPRDERGPNMGQ 591

Qy         61 KLEILIKDTLGLPVAGQTSEFVNQVLEKTAEGNPTGGLVGLRIPTSKV 108
              ||||||||||||||||||||||||||||||||||||||||||||||||||
Db        592 KLEILIKDTLGLPVAGQTSEFVNQVLEKTAEGNPTGGLVGLRIPTSKV 639
```

RESULT 2
US-11-097-143-31914
; Sequence 31914, Application US/11097143
; Publication No. US20050208558A1
; GENERAL INFORMATION:
; APPLICANT: Venter, J. Craig
; APPLICANT: et al.
; TITLE OF INVENTION: DETECTION KIT, SUCH AS NUCLEIC ACID
; TITLE OF INVENTION: ARRAYS, FOR DETECTING EXPRESSION OF 10,000 OR MORE
; TITLE OF INVENTION: DROSOPHILA GENES.
; FILE REFERENCE: CL000728
; CURRENT APPLICATION NUMBER: US/11/097,143
; CURRENT FILING DATE: 2005-04-04
; PRIOR APPLICATION NUMBER: 60/157,832
; PRIOR FILING DATE: 1999-10-05
; PRIOR APPLICATION NUMBER: 60/160,191
; PRIOR FILING DATE: 1999-10-19

```
;  PRIOR APPLICATION NUMBER: 60/161,932
;  PRIOR FILING DATE: 1999-10-28
;  PRIOR APPLICATION NUMBER: 60/164,769
;  PRIOR FILING DATE: 1999-11-12
;  PRIOR APPLICATION NUMBER: 60/173,383
;  PRIOR FILING DATE: 1999-12-28
;  PRIOR APPLICATION NUMBER: 60/175,693
;  PRIOR FILING DATE: 2000-01-12
;  PRIOR APPLICATION NUMBER: 60/184,831
;  PRIOR FILING DATE: 2000-02-24
;  PRIOR APPLICATION NUMBER: 60/191,637
;  PRIOR FILING DATE: 2000-03-23
;  NUMBER OF SEQ ID NOS: 43008
;  SOFTWARE: FastSEQ for Windows Version 4.0
; SEQ ID NO 31914
;   LENGTH: 814
;   TYPE: PRT
;   ORGANISM: DROSOPHILA
US-11-097-143-31914

  Query Match             36.7%;  Score 202;  DB 6;  Length 814;
  Best Local Similarity   43.6%;  Pred. No. 6.6e-14;
  Matches   48;  Conservative   17;  Mismatches   31;  Indels   14;  Gaps    2;

Qy          1 KLAFRACRLKKKAQYEANKVKLWGLNTEYDNLLFVINSIKQEIVNRVQNPRDERGPNMGQ 60
              ||| |||||||||:||||:||:||  |:  |:  |  :|| :| :       |  |:|:
Db        708 KLASRACRLKKKAQHEANKIKLFGLEIEHKRLMNGIAELKQALVVK------HRTKNLGE 761

Qy         61 KLE--------ILIKDTLGLPVAGQTSEFVNQVLEKTAEGNPTGGLVGLR 102
                |       |    : |: :|| :::|||:|||    | | |||  ||
Db        762 STEEVDQQIARIYATASSGIRIAGGSTDFVNKVLENMRGGMPNGGLEELR 811


RESULT 3
US-10-450-763-59634
; Sequence 59634, Application US/10450763
; Publication No. US20050196754A1
; GENERAL INFORMATION:
;  APPLICANT: Hyseq, Inc
;  TITLE OF INVENTION: NOVEL NUCLEIC ACIDS AND POLYPEPTIDES
;  FILE REFERENCE: 790CIP3/US
;  CURRENT APPLICATION NUMBER: US/10/450,763
;  CURRENT FILING DATE:  2003-06-11
;  PRIOR APPLICATION NUMBER: PCT/US01/08631
;  PRIOR FILING DATE: 2001-03-30
;  PRIOR APPLICATION NUMBER: 09/540,217
;  PRIOR FILING DATE: 2000-03-31
;  PRIOR APPLICATION NUMBER: 09/649,167
;  PRIOR FILING DATE: 2000-08-23
;  NUMBER OF SEQ ID NOS: 60736
;  SOFTWARE: Custom
; SEQ ID NO 59634
;   LENGTH: 160
;   TYPE: PRT
;   ORGANISM: Homo sapiens
US-10-450-763-59634

  Query Match             36.4%;  Score 200.5;  DB 5;  Length 160;
  Best Local Similarity   73.0%;  Pred. No. 1.1e-14;
  Matches   46;  Conservative    2;  Mismatches   12;  Indels    3;  Gaps    2;

Qy          9 LKKKAQYEANKVKLW-GLNTEYDNLLFVINSIKQEIVNRVQNPRDERGPNMGQKLEILIK 67
              | |   |: :   ||     |  ||||||||||||||||||||||||||||||||||||
Db          8 LTKNHGYDHS--SLWRSTALEPYNLLFVINSIKQEIVNRVQNPRDERGPNMGQKLEILIK 65

Qy         68 DTL 70
              |||
Db         66 DTL 68


RESULT 4
US-10-282-122A-47201
; Sequence 47201, Application US/10282122A
; Publication No. US20040029129A1
; GENERAL INFORMATION:
;  APPLICANT: Wang, Liangsu
;  APPLICANT:  Zamudio, Carlos
;  APPLICANT:  Malone, Cheryl
;  APPLICANT:  Haselbeck, Robert
;  APPLICANT:  Ohlsen, Kari
;  APPLICANT:  Zyskind, Judith
;  APPLICANT:  Wall, Daniel
;  APPLICANT:  Trawick, John
;  APPLICANT:  Carr, Grant
;  APPLICANT:  Yamamoto, Robert
;  APPLICANT:  Forsyth, R.
;  APPLICANT:  Xu, H.
;  TITLE OF INVENTION: Identification of Essential Genes in Microorganisms
;  FILE REFERENCE: ELITRA.034A
;  CURRENT APPLICATION NUMBER: US/10/282,122A
;  CURRENT FILING DATE:  2003-02-20
;  PRIOR APPLICATION NUMBER: 60/191,078
;  PRIOR FILING DATE: 2000-03-21
;  PRIOR APPLICATION NUMBER: 60/206,848
;  PRIOR FILING DATE: 2000-05-23
;  PRIOR APPLICATION NUMBER: 60/207,727
;  PRIOR FILING DATE: 2000-05-26
;  PRIOR APPLICATION NUMBER: 60/230,335
;  PRIOR FILING DATE: 2000-09-06
;  PRIOR APPLICATION NUMBER: 60/230,347
;  PRIOR FILING DATE: 2000-09-09
;  PRIOR APPLICATION NUMBER: 60/242,578
;  PRIOR FILING DATE: 2000-10-23
;  PRIOR APPLICATION NUMBER: 60/253,625
;  PRIOR FILING DATE: 2000-11-27
;  PRIOR APPLICATION NUMBER: 60/257,931
;  PRIOR FILING DATE: 2000-12-22
;  PRIOR APPLICATION NUMBER: 60/267,636
;  PRIOR FILING DATE: 2001-02-09
;  PRIOR APPLICATION NUMBER: 60/269,308
;  PRIOR FILING DATE: 2001-02-16
;  Remaining Prior Application data removed - See File Wrapper or PALM.
;  NUMBER OF SEQ ID NOS: 78614
;  SOFTWARE: PatentIn version 3.1
; SEQ ID NO 47201
;   LENGTH: 281
;   TYPE: PRT
;   ORGANISM: Borrelia burgdorferi
US-10-282-122A-47201

  Query Match             14.2%;  Score 78.5;  DB 4;  Length 281;
  Best Local Similarity   31.6%;  Pred. No. 2.6;
  Matches   24;  Conservative   14;  Mismatches   23;  Indels   15;  Gaps    4;

Qy         21 KLWGLNTEYDNLLFVIN-SIKQEIVNRVQNPRDER----GPNMGQKLEILIKDT---LGL 72
              |||| |       ::|| ||:|:|:  :    :|:    || :|   |||:| |
Db         31 KLWGQN-------YLINESIRQKIIESLDIKENEKIWEIGPGLGAMTEILLKKTNLLTAF 83

Qy         73 PVAGQTSEFVNQVLEK 88
              :  : || :|:   |
Db         84 EIDLKYSEILNEKFGK 99


RESULT 5
US-10-472-928-2234
; Sequence 2234, Application US/10472928
; Publication No. US20050020813A1
; GENERAL INFORMATION:
;  APPLICANT: CHIRON SpA
;  APPLICANT: THE INSTITUTE FOR GENOMIC RESEARCH
;  TITLE OF INVENTION: STREPTOCOCCUS PNEUMONIAE PROTEINS AND NUCLEIC ACIDS
;  FILE REFERENCE: P026926WO
;  CURRENT APPLICATION NUMBER: US/10/472,928
;  CURRENT FILING DATE:  2003-09-26
;  PRIOR APPLICATION NUMBER: GB-0107658.7
;  PRIOR FILING DATE: 2001-03-27
```

```
;  NUMBER OF SEQ ID NOS: 4979
;  SOFTWARE: SeqWin99, version 1.03
; SEQ ID NO 2234
;   LENGTH: 345
;   TYPE: PRT
;   ORGANISM: Streptococcus pneumoniae
;   FEATURE:
;   OTHER INFORMATION: 1,4-alpha-glucan branching enzyme-related
US-10-472-928-2234

  Query Match             13.7%;  Score 75.5;  DB 5;  Length 345;
  Best Local Similarity   27.6%;  Pred. No. 7.6;
  Matches   29;  Conservative   17;  Mismatches   20;  Indels   39;  Gaps    6;

Qy          9 LKKKAQYEANKVK--------LWGLNTEYDNLLFVINSIKQEIVNRVQNPRDERGPNM-- 58
              :  | :| |:::          || ::| || :          ||::      | |  ::
Db        193 MNAKMKYFASQLNQFYKDHRCLWEIDTSYDGI---------EIID-----ADNRDQSVLS 238

Qy         59 ----GQKLEILI----------KD-TLGLPVAGQTSEFVNQVLEK 88
                  |:| |:|:          || |:|||||   |  |  ||:
Db        239 FIRKGKKGEMLVCIFNMVPVERKDFTIGLPVAGIYEEVWNTELEE 283


RESULT 6
US-10-369-493-22667
; Sequence 22667, Application US/10369493
; Publication No. US20030233675A1
; GENERAL INFORMATION:
;  APPLICANT: Cao, Yongwei
;  APPLICANT:  Hinkle, Gregory J.
;  APPLICANT:  Slater, Steven C.
;  APPLICANT:  Goldman, Barry S.
;  APPLICANT:  Chen, Xianfeng
;  TITLE OF INVENTION: EXPRESSION OF MICROBIAL PROTEINS IN PLANTS FOR PRODUCTION OF
;  TITLE OF INVENTION:  PLANTS WITH IMPROVED PROPERTIES
;  FILE REFERENCE: 38-10(52052)B
;  CURRENT APPLICATION NUMBER: US/10/369,493
;  CURRENT FILING DATE:  2003-02-28
;  PRIOR APPLICATION NUMBER: US 60/360,039
;  PRIOR FILING DATE: 2002-02-21
;  NUMBER OF SEQ ID NOS: 47374
; SEQ ID NO 22667
;   LENGTH: 1175
;   TYPE: PRT
;   ORGANISM: Schizosaccharomyces pombe
;   FEATURE:
;   NAME/KEY: unsure
;   LOCATION: (1)..(1175)
;   OTHER INFORMATION: unsure at all Xaa locations
US-10-369-493-22667

  Query Match             13.7%;  Score 75.5;  DB 4;  Length 1175;
  Best Local Similarity   25.0%;  Pred. No. 39;
  Matches   23;  Conservative   21;  Mismatches   41;  Indels    7;  Gaps    2;

Qy          2 LAFRACRLKKKAQYEANKVKLWGLNTEYDNLLFVINSIKQEIVNRVQNPRDERGPNMGQK 61
              :   :||   : : :  : : :     | |:||:|:::|:  :| |: | ||   : |:
Db         38 VCLNSCRFYNETRSQLREKRFFVGXNLYINILFLISNVKRIRASRNQDNRPERQSRL-QR 96

Qy         62 LEILIKDTLGLPVAGQTSEFVNQVLEKTAEGN 93
                ||:   |       :  | || :| | |
Db         97 SSSLIEQVRG------NEDGENDVLNQTRETN 122


RESULT 7
US-10-369-493-19968
; Sequence 19968, Application US/10369493
; Publication No. US20030233675A1
; GENERAL INFORMATION:
;  APPLICANT: Cao, Yongwei
;  APPLICANT:  Hinkle, Gregory J.
;  APPLICANT:  Slater, Steven C.
;  APPLICANT:  Goldman, Barry S.
;  APPLICANT:  Chen, Xianfeng
;  TITLE OF INVENTION: EXPRESSION OF MICROBIAL PROTEINS IN PLANTS FOR PRODUCTION OF
;  TITLE OF INVENTION:  PLANTS WITH IMPROVED PROPERTIES
;  FILE REFERENCE: 38-10(52052)B
;  CURRENT APPLICATION NUMBER: US/10/369,493
;  CURRENT FILING DATE:  2003-02-28
;  PRIOR APPLICATION NUMBER: US 60/360,039
;  PRIOR FILING DATE: 2002-02-21
;  NUMBER OF SEQ ID NOS: 47374
; SEQ ID NO 19968
;   LENGTH: 1124
;   TYPE: PRT
;   ORGANISM: No. US20030233675A1toc punctiforme
;   FEATURE:
;   NAME/KEY: unsure
;   LOCATION: (1)..(1124)
;   OTHER INFORMATION: unsure at all Xaa locations
US-10-369-493-19968

  Query Match             13.5%;  Score 74.5;  DB 4;  Length 1124;
  Best Local Similarity   28.6%;  Pred. No. 48;
  Matches   26;  Conservative   11;  Mismatches   37;  Indels   17;  Gaps    4;

Qy         24 GLNTEYDNLLFVINSIKQEIVNRVQNPRDERGPNMGQKLEILIKDTLGLPVA-------- 75
              |: |    |: | :| :  ::|     ||| |   : |          | |:|
Db        316 GVETGGSNIQFAVNPLNGDVVVIEMNPRVSRSSALSSK-------ATGFPIAKMAAKLAV 368

Qy         76 GQT-SEFVNQVLEKT-AEGNPTGGLVGLRIP 104
              | |  |  | : :|| |   ||   |  :||
Db        369 GYTLDEIKNDITKKTPASFEPTIDYVVTKIP 399


RESULT 8
US-10-282-122A-71587
; Sequence 71587, Application US/10282122A
; Publication No. US20040029129A1
; GENERAL INFORMATION:
;  APPLICANT: Wang, Liangsu
;  APPLICANT:  Zamudio, Carlos
;  APPLICANT:  Malone, Cheryl
;  APPLICANT:  Haselbeck, Robert
;  APPLICANT:  Ohlsen, Kari
;  APPLICANT:  Zyskind, Judith
;  APPLICANT:  Wall, Daniel
;  APPLICANT:  Trawick, John
;  APPLICANT:  Carr, Grant
;  APPLICANT:  Yamamoto, Robert
;  APPLICANT:  Forsyth, R.
;  APPLICANT:  Xu, H.
;  TITLE OF INVENTION: Identification of Essential Genes in Microorganisms
;  FILE REFERENCE: ELITRA.034A
;  CURRENT APPLICATION NUMBER: US/10/282,122A
;  CURRENT FILING DATE:  2003-02-20
;  PRIOR APPLICATION NUMBER: 60/191,078
;  PRIOR FILING DATE: 2000-03-21
;  PRIOR APPLICATION NUMBER: 60/206,848
;  PRIOR FILING DATE: 2000-05-23
;  PRIOR APPLICATION NUMBER: 60/207,727
;  PRIOR FILING DATE: 2000-05-26
;  PRIOR APPLICATION NUMBER: 60/230,335
;  PRIOR FILING DATE: 2000-09-06
;  PRIOR APPLICATION NUMBER: 60/230,347
;  PRIOR FILING DATE: 2000-09-09
;  PRIOR APPLICATION NUMBER: 60/242,578
;  PRIOR FILING DATE: 2000-10-23
;  PRIOR APPLICATION NUMBER: 60/253,625
;  PRIOR FILING DATE: 2000-11-27
;  PRIOR APPLICATION NUMBER: 60/257,931
;  PRIOR FILING DATE: 2000-12-22
;  PRIOR APPLICATION NUMBER: 60/267,636
;  PRIOR FILING DATE: 2001-02-09
;  PRIOR APPLICATION NUMBER: 60/269,308
;  PRIOR FILING DATE: 2001-02-16
```

```
;  Remaining Prior Application data removed - See File Wrapper or PALM.
;  NUMBER OF SEQ ID NOS: 78614
;  SOFTWARE: PatentIn version 3.1
; SEQ ID NO 71587
;    LENGTH: 625
;    TYPE: PRT
;    ORGANISM: Staphylococcus haemolyticus
;    FEATURE:
;    NAME/KEY: MISC_FEATURE
;    LOCATION: (608)..(608)
;    OTHER INFORMATION: X=any amino acid
US-10-282-122A-71587

  Query Match             13.3%;  Score 73.5;  DB 4;  Length 625;
  Best Local Similarity   21.8%;  Pred. No. 28;
  Matches   22;  Conservative   17;  Mismatches   31;  Indels   31;  Gaps    2;

Qy          8 RLKKKAQYEANKVKLWGLNTEYDNLLF----------------------VINSIKQEI 43
              ||: :  |  :::| :   | | :|||                      ::||: | :
Db        130 RLQTRLSYTESQIKFYKTKTVYTDLLFLYLEKAFLSQNFFDIPMMHSQLDDILNSMFQYL 189

Qy         44 VNRVQNPRDERGPNMGQKLEILIKDTLGLPVAGQTSEFVNQ 84
               |  ||   |   : |::   | | |       | : :|:
Db        190 PNFFQNETSEDNMYLAQRIMFQIDDML-------TKDMLNE 223


RESULT 9
US-10-369-493-19087
; Sequence 19087, Application US/10369493
; Publication No. US20030233675A1
; GENERAL INFORMATION:
;  APPLICANT: Cao, Yongwei
;  APPLICANT:  Hinkle, Gregory J.
;  APPLICANT:  Slater, Steven C.
;  APPLICANT:  Goldman, Barry S.
;  APPLICANT:  Chen, Xianfeng
;  TITLE OF INVENTION: EXPRESSION OF MICROBIAL PROTEINS IN PLANTS FOR PRODUCTION OF
;  TITLE OF INVENTION:  PLANTS WITH IMPROVED PROPERTIES
;  FILE REFERENCE: 38-10(52052)B
;  CURRENT APPLICATION NUMBER: US/10/369,493
;  CURRENT FILING DATE:  2003-02-28
;  PRIOR APPLICATION NUMBER: US 60/360,039
;  PRIOR FILING DATE: 2002-02-21
;  NUMBER OF SEQ ID NOS: 47374
; SEQ ID NO 19087
;    LENGTH: 1104
;    TYPE: PRT
;    ORGANISM: Anabaena PCC7120
US-10-369-493-19087

  Query Match             13.3%;  Score 73.5;  DB 4;  Length 1104;
  Best Local Similarity   27.5%;  Pred. No. 61;
  Matches   25;  Conservative   12;  Mismatches   37;  Indels   17;  Gaps    4;

Qy         24 GLNTEYDNLLFVINSIKQEIVNRVQNPRDERGPNMGQKLEILIKDTLGLPVA-------- 75
              |: |   |: | :|  :  ::|    |||  |   :  |         | |:|
Db        281 GVETGGSNIQFAVNPVNGDVVVIEMNPRVSRSSALSSK-------ATGFPIAKMAAKLAV 333

Qy         76 GQT-SEFVNQVLEKT-AEGNPTGGLVGLRIP 104
              | |  |  | : :|| |   ||   |  ::|
Db        334 GYTLDEIRNDITKKTPASFEPTIDYVVTKVP 364


RESULT 10
US-10-351-750-1
; Sequence 1, Application US/10351750
; Publication No. US20030138836A1
; GENERAL INFORMATION:
;  APPLICANT: THE SALK INSTITUTE FOR BIOLOGICAL STUDIES
;  APPLICANT:  EVANS, RONALD
;  APPLICANT:  CHEN, J.
;  TITLE OF INVENTION: TRANSCRIPTIONAL CO-REPRESSOR THAT INTERACTS WITH NUCLEAR HORMONE
;  TITLE OF INVENTION:  RECEPTORS
;  FILE REFERENCE: SALK1510-2
;  CURRENT APPLICATION NUMBER: US/10/351,750
;  CURRENT FILING DATE:  2003-01-23
;  PRIOR APPLICATION NUMBER: US/09/337,384
;  PRIOR FILING DATE: 1999-06-21
;  PRIOR APPLICATION NUMBER: 08/522,726
;  PRIOR FILING DATE: 1995-09-01
;  NUMBER OF SEQ ID NOS: 3
;  SOFTWARE: PatentIn version 3.0
; SEQ ID NO 1
;    LENGTH: 1495
;    TYPE: PRT
;    ORGANISM: Homo sapiens
US-10-351-750-1

  Query Match             13.3%;  Score 73.5;  DB 4;  Length 1495;
  Best Local Similarity   30.6%;  Pred. No. 91;
  Matches   22;  Conservative   14;  Mismatches   17;  Indels   19;  Gaps    3;

Qy         37 NSIKQEIVNRVQNPRDERGPNMGQKLEILIKDTLGLPVAGQTSEFVNQVLEKTAEGNPTG 96
              : :||| :    :|| : ||           ::||:| | : |     ||  || |:  |
Db        143 SGVKQEQL----SPRGQAGP----------PESLGVPTAQEAS-----VLRGTALGSVPG 183

Qy         97 GLVGLRIPTSKV 108
              | :   ||:::|
Db        184 GSITKGIPSTRV 195


RESULT 11
US-10-705-165-26
; Sequence 26, Application US/10705165
; Publication No. US20050202440A1
; GENERAL INFORMATION:
;  APPLICANT: Fletterick, Robert
;  APPLICANT:  Hur, Eugene
;  APPLICANT:  Bueher, Ben
;  TITLE OF INVENTION: Inhibitors for Androgen Antagonist Refractory Prostate Cancer
;  FILE REFERENCE: 061040-0018-US
;  CURRENT APPLICATION NUMBER: US/10/705,165
;  CURRENT FILING DATE:  2003-11-06
;  PRIOR APPLICATION NUMBER: US 09/281,717
;  PRIOR FILING DATE: 1999-03-30
;  PRIOR APPLICATION NUMBER: US 60/079,956
;  PRIOR FILING DATE: 1998-03-30
;  PRIOR APPLICATION NUMBER: US 60/113,146
;  PRIOR FILING DATE: 1998-12-16
;  PRIOR APPLICATION NUMBER: US 09/609,361
;  PRIOR FILING DATE: 2000-06-30
;  PRIOR APPLICATION NUMBER: US 09/830,693
;  PRIOR FILING DATE: 1999-03-30
;  PRIOR APPLICATION NUMBER: US 60/113,014
;  PRIOR FILING DATE: 1998-12-16
;  NUMBER OF SEQ ID NOS: 73
;  SOFTWARE: PatentIn version 3.3
; SEQ ID NO 26
;    LENGTH: 1495
;    TYPE: PRT
;    ORGANISM: Homo sapiens
US-10-705-165-26

  Query Match             13.3%;  Score 73.5;  DB 5;  Length 1495;
  Best Local Similarity   30.6%;  Pred. No. 91;
  Matches   22;  Conservative   14;  Mismatches   17;  Indels   19;  Gaps    3;

Qy         37 NSIKQEIVNRVQNPRDERGPNMGQKLEILIKDTLGLPVAGQTSEFVNQVLEKTAEGNPTG 96
              : :||| :    :|| : ||           ::||:| | : |     ||  || |:  |
Db        143 SGVKQEQL----SPRGQAGP----------PESLGVPTAQEAS-----VLRGTALGSVPG 183

Qy         97 GLVGLRIPTSKV 108
              | :   ||:::|
Db        184 GSITKGIPSTRV 195
```

```
RESULT 12
US-09-819-104A-2
; Sequence 2, Application US/09819104A
; Publication No. US20030027137A1
; GENERAL INFORMATION:
;  APPLICANT: Chen, J. Don
;  TITLE OF INVENTION: NOVEL NUCLEAR RECEPTOR COREPRESSOR MOLECULES
;  TITLE OF INVENTION:  AND USES THEREFOR
;  FILE REFERENCE: UMG-030
;  CURRENT APPLICATION NUMBER: US/09/819,104A
;  CURRENT FILING DATE:  2001-03-27
;  PRIOR APPLICATION NUMBER: 60/193,138
;  PRIOR FILING DATE: 2000-03-29
;  NUMBER OF SEQ ID NOS: 6
;  SOFTWARE: PatentIn Ver. 2.0
; SEQ ID NO 2
;   LENGTH: 2507
;   TYPE: PRT
;   ORGANISM: Homo sapiens
US-09-819-104A-2

  Query Match             13.3%;  Score 73.5;  DB 3;  Length 2507;
  Best Local Similarity   30.6%;  Pred. No. 1.8e+02;
  Matches   22;  Conservative   14;  Mismatches   17;  Indels   19;  Gaps    3;

Qy         37 NSIKQEIVNRVQNPRDERGPNMGQKLEILIKDTLGLPVAGQTSEFVNQVLEKTAEGNPTG 96
              : :||| :    :|| : ||          ::||:| | : |     ||  || |: |
Db       1155 SGVKQEQL----SPRGQAGP----------PESLGVPTAQEAS-----VLRGTALGSVPG 1195

Qy         97 GLVGLRIPTSKV 108
              | :   ||:::|
Db       1196 GSITKGIPSTRV 1207


RESULT 13
US-10-723-860-1740
; Sequence 1740, Application US/10723860
; Publication No. US20040253606A1
; GENERAL INFORMATION:
;  APPLICANT: Aziz, Natasha
;  APPLICANT:  Ginsburg, Wendy M.
;  APPLICANT:  Zlotnik, Albert
;  TITLE OF INVENTION: Methods of Diagnosis of Soft Tissue Sarcoma, Compositions &
;  TITLE OF INVENTION:  Methods for Screening for Soft Tissue Sarcoma Modulators
;  FILE REFERENCE: 05882.0193.NPUS01
;  CURRENT APPLICATION NUMBER: US/10/723,860
;  CURRENT FILING DATE:  2003-11-26
;  PRIOR APPLICATION NUMBER: 60/429,739
;  PRIOR FILING DATE: 2002-11-26
;  NUMBER OF SEQ ID NOS: 8393
;  SOFTWARE: PatentIn version 3.2
; SEQ ID NO 1740
;   LENGTH: 2507
;   TYPE: PRT
;   ORGANISM: Homo sapiens
US-10-723-860-1740

  Query Match             13.3%;  Score 73.5;  DB 5;  Length 2507;
  Best Local Similarity   30.6%;  Pred. No. 1.8e+02;
  Matches   22;  Conservative   14;  Mismatches   17;  Indels   19;  Gaps    3;

Qy         37 NSIKQEIVNRVQNPRDERGPNMGQKLEILIKDTLGLPVAGQTSEFVNQVLEKTAEGNPTG 96
              : :||| :    :|| : ||          ::||:| | : |     ||  || |: |
Db       1155 SGVKQEQL----SPRGQAGP----------PESLGVPTAQEAS-----VLRGTALGSVPG 1195

Qy         97 GLVGLRIPTSKV 108
              | :   ||:::|
Db       1196 GSITKGIPSTRV 1207


RESULT 14
US-10-087-192-654
; Sequence 654, Application US/10087192
; Publication No. US20020182586A1
; GENERAL INFORMATION:
;  APPLICANT: Morris, David W.
;  APPLICANT:  Engelhard, Eric K.
;  TITLE OF INVENTION: NOVEL COMPOSITIONS AND METHODS FOR
;  TITLE OF INVENTION:  CANCER
;  FILE REFERENCE: 529452000122
;  CURRENT APPLICATION NUMBER: US/10/087,192
;  CURRENT FILING DATE:  2002-03-01
;  PRIOR APPLICATION NUMBER: US 09/747,377
;  PRIOR FILING DATE: 2000-12-22
;  PRIOR APPLICATION NUMBER: US 09/798,586
;  PRIOR FILING DATE: 2001-03-02
;  NUMBER OF SEQ ID NOS: 2059
;  SOFTWARE: FastSEQ for Windows Version 4.0
; SEQ ID NO 654
;   LENGTH: 2517
;   TYPE: PRT
;   ORGANISM: Homo sapiens
US-10-087-192-654

  Query Match             13.3%;  Score 73.5;  DB 4;  Length 2517;
  Best Local Similarity   30.6%;  Pred. No. 1.8e+02;
  Matches   22;  Conservative   14;  Mismatches   17;  Indels   19;  Gaps    3;

Qy         37 NSIKQEIVNRVQNPRDERGPNMGQKLEILIKDTLGLPVAGQTSEFVNQVLEKTAEGNPTG 96
              : :||| :    :|| : ||          ::||:| | : |     ||  || |: |
Db       1165 SGVKQEQL----SPRGQAGP----------PESLGVPTAQEAS-----VLRGTALGSVPG 1205

Qy         97 GLVGLRIPTSKV 108
              | :   ||:::|
Db       1206 GSITKGIPSTRV 1217


RESULT 15
US-10-617-320-2845
; Sequence 2845, Application US/10617320
; Publication No. US20050136404A1
;  GENERAL INFORMATION:
;       APPLICANT: Lynn A Doucette-Stamm  and David Bush
;       TITLE OF INVENTION: NUCLEIC ACID AND AMINO ACID
;                           SEQUENCES RELATING TO STREPTOCOCCUS PNEUMONIAE   FOR DIAGNOS
;                           THERAPEUTICS
;       NUMBER OF SEQUENCES: 5206
;       CORRESPONDENCE ADDRESS:
;            ADDRESSEE: GENOME THERAPEUTICS CORPORATION
;            STREET: 100 Beaver Street
;            CITY: Waltham
;            STATE: Massachusetts
;            COUNTRY: USA
;            ZIP: 02354
;       COMPUTER READABLE FORM:
;            MEDIUM TYPE: CD/ROM ISO9660
;            COMPUTER: <Unknown>
;            OPERATING SYSTEM: <Unknown>
;            SOFTWARE: <Unknown>
;       CURRENT APPLICATION DATA:
;            APPLICATION NUMBER: US/10/617,320
;            FILING DATE: 10-Jul-2003
;       PRIOR APPLICATION DATA:
;            APPLICATION NUMBER: US/09/107,433
;            FILING DATE: 30-Jun-1998
;            APPLICATION NUMBER: 60/ 085131
;            FILING DATE: May 12, 1998
;            APPLICATION NUMBER: 60/051553
;            FILING DATE: July 2, 1997
;       ATTORNEY/AGENT INFORMATION:
;            NAME: Ariniello, Pamela Deneke
;            REGISTRATION NUMBER: 40,489
;            REFERENCE/DOCKET NUMBER: GTC-011
;       TELECOMMUNICATION INFORMATION:
;            TELEPHONE: (781)893-5007
;            TELEFAX: (781)893-8277
```

```
;   INFORMATION FOR SEQ ID NO: 2845:
;        SEQUENCE CHARACTERISTICS:
;             LENGTH: 651 amino acids
;             TYPE: amino acid
;             TOPOLOGY: linear
;        MOLECULE TYPE: protein
;        HYPOTHETICAL: YES
;        ORIGINAL SOURCE:
;             ORGANISM: Streptococcus pneumoniae
;        FEATURE:
;             NAME/KEY:  misc_feature
;             LOCATION: (B) LOCATION 1...651
;        SEQUENCE DESCRIPTION: SEQ ID NO: 2845:
US-10-617-320-2845

  Query Match             13.2%;  Score 72.5;  DB 5;  Length 651;
  Best Local Similarity   31.0%;  Pred. No. 39;
  Matches   26;  Conservative   12;  Mismatches   15;  Indels   31;  Gaps    5;

Qy         22 LWGLNTEYDNLLFVINSIKQEIVNRVQNPRDERGPNM------GQKLEILI--------- 66
              || ::| || :          ||::      | |  ::      |:| |:|:
Db        520 LWEIDTSYDGI---------EIID-----ADNRDQSVLSFIRKGKKGEMLVCIFNMVPVE 565

Qy         67 -KD-TLGLPVAGQTSEFVNQVLEK 88
               || |:||||||   |  |  ||:
Db        566 RKDFTIGLPVAGIYEEVWNTELEE 589


Search completed: February 28, 2006, 09:02:04
Job time : 32.4887 secs
```

GenCore version 5.1.7

OM protein - protein search, using sw model

Run on: February 28, 2006, 08:56:26 ; Search time 2.88722 Seconds
(without alignments)
556.876 Million cell updates/sec

Title: US-10-717-665A-44_COPY_532_639
Perfect score: 551
Sequence: 1 KLAFRACRLKKKAQYEANKV.....TAEGNPTGGLVGLRIPTSKV 108

Scoring table: BLOSUM62
Gapop 10.0 , Gapext 0.5

Searched: 117670 seqs, 14887254 residues

Total number of hits satisfying chosen parameters: 117670

Minimum DB seq length: 0
Maximum DB seq length: 2000000000

Post-processing: Minimum Match 0%
Maximum Match 100%
Listing first 45 summaries

Database : Published_Applications_AA_New:*
1: /cgn2_6/ptodata/2/pubpaa/US08_NEW_PUB.pep:*
2: /cgn2_6/ptodata/2/pubpaa/US06_NEW_PUB.pep:*
3: /cgn2_6/ptodata/2/pubpaa/US07_NEW_PUB.pep:*
4: /cgn2_6/ptodata/2/pubpaa/PCT_NEW_PUB.pep:*
5: /cgn2_6/ptodata/2/pubpaa/US09_NEW_PUB.pep:*
6: /cgn2_6/ptodata/2/pubpaa/US10_NEW_PUB.pep:*
7: /cgn2_6/ptodata/2/pubpaa/US11_NEW_PUB.pep:*
8: /cgn2_6/ptodata/2/pubpaa/US60_NEW_PUB.pep:*

Pred. No. is the number of results predicted by chance to have a score greater than or equal to the score of the result being printed, and is derived by analysis of the total score distribution.

SUMMARIES

Result No.	Score	% Query Match	Length	DB	ID	Description
1	69	12.5	803	6	US-10-821-234-1643	Sequence 1643, Ap
2	69	12.5	803	6	US-10-962-951-2	Sequence 2, Appli
3	67	12.2	1410	6	US-10-878-556A-136	Sequence 136, App
4	67	12.2	1586	6	US-10-821-234-901	Sequence 901, App
5	66	12.0	207	6	US-10-467-657-314	Sequence 314, App
6	66	12.0	207	6	US-10-467-657-3272	Sequence 3272, Ap
7	65.5	11.9	313	7	US-11-087-719-15	Sequence 15, Appl
8	65.5	11.9	323	7	US-11-087-719-14	Sequence 14, Appl
9	65.5	11.9	338	7	US-11-087-719-13	Sequence 13, Appl
10	65.5	11.9	339	6	US-10-821-234-1507	Sequence 1507, Ap
11	65.5	11.9	339	7	US-11-072-175-143	Sequence 143, App
12	65.5	11.9	378	7	US-11-207-626A-34	Sequence 34, Appl
13	65.5	11.9	1066	7	US-11-055-822-370	Sequence 370, App
14	65.5	11.9	1066	7	US-11-055-822-1002	Sequence 1002, Ap
15	65.5	11.9	1113	7	US-11-055-822-368	Sequence 368, App
16	65.5	11.9	1113	7	US-11-055-822-1000	Sequence 1000, Ap
17	65	11.8	1036	7	US-11-072-512-2812	Sequence 2812, Ap
18	64.5	11.7	71	6	US-10-962-951-14	Sequence 14, Appl
19	64.5	11.7	71	6	US-10-962-951-15	Sequence 15, Appl
20	64.5	11.7	71	6	US-10-962-951-16	Sequence 16, Appl
21	64.5	11.7	71	6	US-10-962-951-17	Sequence 17, Appl
22	62.5	11.3	343	7	US-11-072-512-3068	Sequence 3068, Ap
23	62	11.3	2710	7	US-11-051-453-41	Sequence 41, Appl
24	61.5	11.2	346	7	US-11-072-175-142	Sequence 142, App
25	61.5	11.2	770	6	US-10-821-234-1269	Sequence 1269, Ap
26	61	11.1	333	6	US-10-821-234-1323	Sequence 1323, Ap
27	61	11.1	333	6	US-10-878-556A-85	Sequence 85, Appl
28	60	10.9	504	7	US-11-121-438-6	Sequence 6, Appli
29	60	10.9	636	6	US-10-485-517-170	Sequence 170, App
30	59	10.7	509	7	US-11-024-959-393	Sequence 393, App
31	59	10.7	793	7	US-11-060-914-2	Sequence 2, Appli
32	59	10.7	831	7	US-11-198-746-5	Sequence 5, Appli
33	59	10.7	831	7	US-11-198-794-5	Sequence 5, Appli
34	58.5	10.6	281	7	US-11-212-443-180	Sequence 180, App
35	58.5	10.6	284	7	US-11-212-443-58	Sequence 58, Appl
36	58.5	10.6	338	6	US-10-878-556A-19	Sequence 19, Appl
37	58.5	10.6	1574	7	US-11-212-443-179	Sequence 179, App
38	58.5	10.6	1978	7	US-11-212-443-60	Sequence 60, Appl
39	58	10.5	319	7	US-11-120-308-80	Sequence 80, Appl
40	58	10.5	331	7	US-11-202-566-28	Sequence 28, Appl
41	58	10.5	355	7	US-11-237-600-5	Sequence 5, Appli
42	58	10.5	597	6	US-10-467-657-5218	Sequence 5218, Ap
43	58	10.5	700	7	US-11-196-475-66	Sequence 66, Appl
44	58	10.5	833	7	US-11-198-746-8	Sequence 8, Appli
45	58	10.5	833	7	US-11-198-794-8	Sequence 8, Appli

ALIGNMENTS

RESULT 1
US-10-821-234-1643
; Sequence 1643, Application US/10821234
; Publication No. US20050255114A1
; GENERAL INFORMATION:
; APPLICANT: Labat, Ivan
; APPLICANT: Stache-Crain, Birgit
; APPLICANT: Andarmani, Susan
; APPLICANT: Tang, Y. Tom
; TITLE OF INVENTION: Methods for Diagnosis and Treatment of Preeclampsia
; FILE REFERENCE: 821A
; CURRENT APPLICATION NUMBER: US/10/821,234
; CURRENT FILING DATE: 2004-04-07
; PRIOR APPLICATION NUMBER: US 60/462,047
; PRIOR FILING DATE: 2003-04-07
; NUMBER OF SEQ ID NOS: 1704
; SOFTWARE: pt_SEQ_genes Version 1.0
; SEQ ID NO 1643
; LENGTH: 803
; TYPE: PRT
; ORGANISM: Homo sapiens
US-10-821-234-1643

Query Match 12.5%; Score 69; DB 6; Length 803;
Best Local Similarity 27.3%; Pred. No. 7.3;
Matches 24; Conservative 19; Mismatches 33; Indels 12; Gaps 4;

```
Qy          18 NKVKLWGLNTEYDNLLFVINSIKQEIVNRVQNPRDER-----GPNMGQKLEILIKDTLGL 72
               :|::|  |  |  | |     |  :|:  :::  :::         :|   | |:|: ||
Db         113 DKIRLISLTDE--NAL----SGNEELTVKIKCDKEKNLLHVTDTGVGMTREELVKN-LGT 165

Qy          73 PVAGQTSEFVNQVLEKTAEGNPTGGLVG 100
                  ||||:|:: |   :|  |  |:|
Db         166 IAKSGTSEFLNKMTEAQEDGQSTSELIG 193
```

RESULT 2
US-10-962-951-2
; Sequence 2, Application US/10962951
; Publication No. US20060029610A1
; GENERAL INFORMATION:
; APPLICANT: Argon, Yair
; APPLICANT: Gidalevitz, Tali
; APPLICANT: Biswas, Chhanda
; APPLICANT: Simen, Birgitte B.
; APPLICANT: Wanderling, Sherry
; APPLICANT: Ostrovsky, Olga
; TITLE OF INVENTION: GRP94-BASED COMPOSITIONS AND METHODS OF

```
;  TITLE OF INVENTION:  USE THEREOF
;  FILE REFERENCE: 3460-CHOP.C-206US
;  CURRENT APPLICATION NUMBER: US/10/962,951
;  CURRENT FILING DATE:  2004-10-12
;  PRIOR APPLICATION NUMBER: US/10/844,711
;  PRIOR FILING DATE: 2004-05-12
;  PRIOR APPLICATION NUMBER: 60/469,723
;  PRIOR FILING DATE: 2003-05-12
;  PRIOR APPLICATION NUMBER: 60/477,990
;  PRIOR FILING DATE: 2003-06-12
;  PRIOR APPLICATION NUMBER: 60/478,149
;  PRIOR FILING DATE: 2003-06-12
;  PRIOR APPLICATION NUMBER: 60/556,362
;  PRIOR FILING DATE: 2004-04-28
;  PRIOR APPLICATION NUMBER: 60/566,363
;  PRIOR FILING DATE: 2004-04-28
;  NUMBER OF SEQ ID NOS: 27
;  SOFTWARE: FastSEQ for Windows Version 3.0
; SEQ ID NO 2
;   LENGTH: 803
;   TYPE: PRT
;   ORGANISM: Artificial Sequence
;   FEATURE:
;   OTHER INFORMATION: amino acid sequence human GRP94
US-10-962-951-2

  Query Match             12.5%;  Score 69;  DB 6;  Length 803;
  Best Local Similarity   27.3%;  Pred. No. 7.3;
  Matches   24;  Conservative   19;  Mismatches   33;  Indels   12;  Gaps    4;

Qy          18 NKVKLWGLNTEYDNLLFVINSIKQEIVNRVQNPRDER-----GPNMGQKLEILIKDTLGL 72
               :|::|  |  |  | |    |  :|:  :::  :::         :|   | |:|: ||
Db         113 DKIRLISLTDE--NAL----SGNEELTVKIKCDKEKNLLHVTDTGVGMTREELVKN-LGT 165

Qy          73 PVAGQTSEFVNQVLEKTAEGNPTGGLVG 100
                  ||||:|:: |   :|  |  |:|
Db         166 IAKSGTSEFLNKMTEAQEDGQSTSELIG 193


RESULT 3
US-10-878-556A-136
; Sequence 136, Application US/10878556A
; Publication No. US20050266399A1
; GENERAL INFORMATION:
;  APPLICANT: Hoffmann La-Roche Inc.
;  TITLE OF INVENTION: HCV regulated protein expression
;  FILE REFERENCE: 21762
;  CURRENT APPLICATION NUMBER: US/10/878,556A
;  CURRENT FILING DATE:  2004-06-28
;  NUMBER OF SEQ ID NOS: 199
;  SOFTWARE: PatentIn version 3.1
; SEQ ID NO 136
;   LENGTH: 1410
;   TYPE: PRT
;   ORGANISM: Homo sapiens
;  PUBLICATION INFORMATION:
;   DATABASE ACCESSION NUMBER: sw_hum/rrb1_human
;   DATABASE ENTRY DATE: 2003-02-28
US-10-878-556A-136

  Query Match             12.2%;  Score 67;  DB 6;  Length 1410;
  Best Local Similarity   33.9%;  Pred. No. 26;
  Matches   20;  Conservative    9;  Mismatches   14;  Indels   16;  Gaps    3;

Qy          48 QNPRDERGPNMGQKLEILIKDTLGLPVAGQTSEFV-NQVLE--------KTAEGNPTGG 97
               |  : || || |:| |       | |: |: :: | ||  :         | |||:|: |
Db         526 QGKKAERSPNQGKKGE-------GAPIQGKKADSVANQGTKVEGITNQGKKAEGSPSEG 577


RESULT 4
US-10-821-234-901
; Sequence 901, Application US/10821234
; Publication No. US20050255114A1
; GENERAL INFORMATION:
;  APPLICANT: Labat, Ivan
;  APPLICANT:  Stache-Crain, Birgit
;  APPLICANT:  Andarmani, Susan
;  APPLICANT:  Tang, Y. Tom
;  TITLE OF INVENTION: Methods for Diagnosis and Treatment of Preeclampsia
;  FILE REFERENCE: 821A
;  CURRENT APPLICATION NUMBER: US/10/821,234
;  CURRENT FILING DATE:  2004-04-07
;  PRIOR APPLICATION NUMBER: US 60/462,047
;  PRIOR FILING DATE: 2003-04-07
;  NUMBER OF SEQ ID NOS: 1704
;  SOFTWARE: pt_SEQ_genes Version 1.0
; SEQ ID NO 901
;   LENGTH: 1586
;   TYPE: PRT
;   ORGANISM: Homo sapiens
US-10-821-234-901

  Query Match             12.2%;  Score 67;  DB 6;  Length 1586;
  Best Local Similarity   33.9%;  Pred. No. 30;
  Matches   20;  Conservative    9;  Mismatches   14;  Indels   16;  Gaps    3;

Qy          48 QNPRDERGPNMGQKLEILIKDTLGLPVAGQTSEFV-NQVLE--------KTAEGNPTGG 97
               |  : || || |:| |       | |: |: :: | ||  :         | |||:|: |
Db         702 QGKKAERSPNQGKKGE-------GAPIQGKKADSVANQGTKVEGITNQGKKAEGSPSEG 753


RESULT 5
US-10-467-657-314
; Sequence 314, Application US/10467657
; Publication No. US20050260581A1
; GENERAL INFORMATION:
;  APPLICANT: CHIRON SpA
;  APPLICANT: FONTANA Maria Rita
;  APPLICANT: PIZZA Mariagrazia
;  APPLICANT: MASIGNANI Vega
;  APPLICANT: MONACI Elisabetta
;  TITLE OF INVENTION: GONOCOCCAL PROTEINS AND NUCLEIC ACIDS
;  FILE REFERENCE:
;  CURRENT APPLICATION NUMBER: US/10/467,657
;  CURRENT FILING DATE:  2003-08-11
;  PRIOR APPLICATION NUMBER: GB-0103424.8
;  PRIOR FILING DATE: 2001-02-12
;  NUMBER OF SEQ ID NOS: 9218
;  SOFTWARE: SeqWin99, version 1.04
; SEQ ID NO 314
;   LENGTH: 207
;   TYPE: PRT
;   ORGANISM: Neisseria gonorrhoeae
US-10-467-657-314

  Query Match             12.0%;  Score 66;  DB 6;  Length 207;
  Best Local Similarity   22.8%;  Pred. No. 2.6;
  Matches   26;  Conservative   17;  Mismatches   45;  Indels   26;  Gaps    4;

Qy           7 CRLK--------KKAQYEANKVKLWGLNTE----YDNLLFVINSIKQEIVNRVQNPRDER 54
               ||||        : |:|  |::   ||       ||||           ||
Db          14 CRLKNVVWKTLGEVAEYSKNRICSDKLNEHNYVGVDNLL----------QNREGKKLSGY 63

Qy          55 GPNMGQKLEILIKDTLGLPVAGQTSEFVNQVLEKTAEGNPTGGLVGLRIPTSKV 108
               |: |:  | :: | |     |    :: :: :    |   | :: :|:   ||
Db          64 VPSEGKMTEYIVNDIL----IGNIRPYLKKIWQADCTGGTNGDVLVIRVTDEKV 113


RESULT 6
US-10-467-657-3272
; Sequence 3272, Application US/10467657
; Publication No. US20050260581A1
; GENERAL INFORMATION:
;  APPLICANT: CHIRON SpA
;  APPLICANT: FONTANA Maria Rita
;  APPLICANT: PIZZA Mariagrazia
```

```
;  APPLICANT: MASIGNANI Vega
;  APPLICANT: MONACI Elisabetta
;  TITLE OF INVENTION: GONOCOCCAL PROTEINS AND NUCLEIC ACIDS
;  FILE REFERENCE:
;  CURRENT APPLICATION NUMBER: US/10/467,657
;  CURRENT FILING DATE:  2003-08-11
;  PRIOR APPLICATION NUMBER: GB-0103424.8
;  PRIOR FILING DATE: 2001-02-12
;  NUMBER OF SEQ ID NOS: 9218
;  SOFTWARE: SeqWin99, version 1.04
; SEQ ID NO 3272
;   LENGTH: 207
;   TYPE: PRT
;   ORGANISM: Neisseria gonorrhoeae
US-10-467-657-3272

  Query Match             12.0%;  Score 66;  DB 6;  Length 207;
  Best Local Similarity   22.8%;  Pred. No. 2.6;
  Matches   26;  Conservative   17;  Mismatches   45;  Indels   26;  Gaps    4;

Qy          7 CRLK--------KKAQYEANKVKLWGLNTE----YDNLLFVINSIKQEIVNRVQNPRDER 54
              ||||        : |:|  |::    ||       ||||  :         ||
Db         14 CRLKNVVWKTLGEVAEYSKNRICSDKLNEHNYVGVDNLL----------QNREGKKLSGY 63

Qy         55 GPNMGQKLEILIKDTLGLPVAGQTSEFVNQVLEKTAEGNPTGGLVGLRIPTSKV 108
               |: |:  | :: | |     |    :: :: :    |   | :: :|:   ||
Db         64 VPSEGKMTEYIVNDIL----IGNIRPYLKKIWQADCTGGTNGDVLVIRVTDEKV 113


RESULT 7
US-11-087-719-15
; Sequence 15, Application US/11087719
; Publication No. US20050277575A1
; GENERAL INFORMATION:
;  APPLICANT: Semov, Alexander
;  APPLICANT:  Onichtchenko, Anatoli
;  APPLICANT:  Iourtchenko, Ludmila
;  APPLICANT:  Ochiette, Benoit
;  APPLICANT:  Pietrzynski, Grzegorz
;  APPLICANT:  Alakhov, Valery
;  TITLE OF INVENTION: THERAPEUTIC COMPOSITIONS AND METHODS FOR TREATING DISEASES THAT
;  TITLE OF INVENTION:  INVOLVE ANGIOGENESIS
;  FILE REFERENCE: P08605US01/BAS
;  CURRENT APPLICATION NUMBER: US/11/087,719
;  CURRENT FILING DATE:  2005-03-24
;  PRIOR APPLICATION NUMBER: US 60/555,668
;  PRIOR FILING DATE: 2004-03-24
;  NUMBER OF SEQ ID NOS: 88
;  SOFTWARE: PatentIn version 3.3
; SEQ ID NO 15
;   LENGTH: 313
;   TYPE: PRT
;   ORGANISM: Homo sapiens
US-11-087-719-15

  Query Match             11.9%;  Score 65.5;  DB 7;  Length 313;
  Best Local Similarity   43.2%;  Pred. No. 5.2;
  Matches   19;  Conservative   11;  Mismatches    9;  Indels    5;  Gaps    3;

Qy          9 LKKKAQYEANKVK--LWGLNTEYDNLLFVINS-IKQEI--VNRV 47
              ||  |||:|:::|  : || |: |:|: :| |   ||:  :|||
Db         77 LKTPAQYDASELKASMKGLGTDEDSLIEIICSRTNQELQEINRV 120


RESULT 8
US-11-087-719-14
; Sequence 14, Application US/11087719
; Publication No. US20050277575A1
; GENERAL INFORMATION:
;  APPLICANT: Semov, Alexander
;  APPLICANT:  Onichtchenko, Anatoli
;  APPLICANT:  Iourtchenko, Ludmila
;  APPLICANT:  Ochiette, Benoit
;  APPLICANT:  Pietrzynski, Grzegorz
;  APPLICANT:  Alakhov, Valery
;  TITLE OF INVENTION: THERAPEUTIC COMPOSITIONS AND METHODS FOR TREATING DISEASES THAT
;  TITLE OF INVENTION:  INVOLVE ANGIOGENESIS
;  FILE REFERENCE: P08605US01/BAS
;  CURRENT APPLICATION NUMBER: US/11/087,719
;  CURRENT FILING DATE:  2005-03-24
;  PRIOR APPLICATION NUMBER: US 60/555,668
;  PRIOR FILING DATE: 2004-03-24
;  NUMBER OF SEQ ID NOS: 88
;  SOFTWARE: PatentIn version 3.3
; SEQ ID NO 14
;   LENGTH: 323
;   TYPE: PRT
;   ORGANISM: Homo sapiens
US-11-087-719-14

  Query Match             11.9%;  Score 65.5;  DB 7;  Length 323;
  Best Local Similarity   43.2%;  Pred. No. 5.4;
  Matches   19;  Conservative   11;  Mismatches    9;  Indels    5;  Gaps    3;

Qy          9 LKKKAQYEANKVK--LWGLNTEYDNLLFVINS-IKQEI--VNRV 47
              ||  |||:|:::|  : || |: |:|: :| |   ||:  :|||
Db         87 LKTPAQYDASELKASMKGLGTDEDSLIEIICSRTNQELQEINRV 130


RESULT 9
US-11-087-719-13
; Sequence 13, Application US/11087719
; Publication No. US20050277575A1
; GENERAL INFORMATION:
;  APPLICANT: Semov, Alexander
;  APPLICANT:  Onichtchenko, Anatoli
;  APPLICANT:  Iourtchenko, Ludmila
;  APPLICANT:  Ochiette, Benoit
;  APPLICANT:  Pietrzynski, Grzegorz
;  APPLICANT:  Alakhov, Valery
;  TITLE OF INVENTION: THERAPEUTIC COMPOSITIONS AND METHODS FOR TREATING DISEASES THAT
;  TITLE OF INVENTION:  INVOLVE ANGIOGENESIS
;  FILE REFERENCE: P08605US01/BAS
;  CURRENT APPLICATION NUMBER: US/11/087,719
;  CURRENT FILING DATE:  2005-03-24
;  PRIOR APPLICATION NUMBER: US 60/555,668
;  PRIOR FILING DATE: 2004-03-24
;  NUMBER OF SEQ ID NOS: 88
;  SOFTWARE: PatentIn version 3.3
; SEQ ID NO 13
;   LENGTH: 338
;   TYPE: PRT
;   ORGANISM: Homo sapiens
US-11-087-719-13

  Query Match             11.9%;  Score 65.5;  DB 7;  Length 338;
  Best Local Similarity   43.2%;  Pred. No. 5.8;
  Matches   19;  Conservative   11;  Mismatches    9;  Indels    5;  Gaps    3;

Qy          9 LKKKAQYEANKVK--LWGLNTEYDNLLFVINS-IKQEI--VNRV 47
              ||  |||:|:::|  : || |: |:|: :| |   ||:  :|||
Db        102 LKTPAQYDASELKASMKGLGTDEDSLIEIICSRTNQELQEINRV 145


RESULT 10
US-10-821-234-1507
; Sequence 1507, Application US/10821234
; Publication No. US20050255114A1
; GENERAL INFORMATION:
;  APPLICANT: Labat, Ivan
;  APPLICANT:  Stache-Crain, Birgit
;  APPLICANT:  Andarmani, Susan
;  APPLICANT:  Tang, Y. Tom
;  TITLE OF INVENTION: Methods for Diagnosis and Treatment of Preeclampsia
;  FILE REFERENCE: 821A
;  CURRENT APPLICATION NUMBER: US/10/821,234
```

```
;  CURRENT FILING DATE:  2004-04-07
;  PRIOR APPLICATION NUMBER: US 60/462,047
;  PRIOR FILING DATE: 2003-04-07
;  NUMBER OF SEQ ID NOS: 1704
;  SOFTWARE: pt_SEQ_genes Version 1.0
; SEQ ID NO 1507
;    LENGTH: 339
;    TYPE: PRT
;    ORGANISM: Homo sapiens
US-10-821-234-1507

  Query Match             11.9%;  Score 65.5;  DB 6;  Length 339;
  Best Local Similarity   43.2%;  Pred. No. 5.8;
  Matches   19;  Conservative   11;  Mismatches    9;  Indels    5;  Gaps    3;

Qy          9 LKKKAQYEANKVK--LWGLNTEYDNLLFVINS-IKQEI--VNRV 47
              ||  |||:|:::|  : || |: |:|: :| |   ||:  :|||
Db        103 LKTPAQYDASELKASMKGLGTDEDSLIEIICSRTNQELQEINRV 146


RESULT 11
US-11-072-175-143
; Sequence 143, Application US/11072175
; Publication No. US20060029944A1
; GENERAL INFORMATION:
;  APPLICANT: Bristol-Myers Squibb Company
;  TITLE OF INVENTION: IDENTIFICATION OF GENES FOR PREDICTING ACTIVITY OF COMPOUNDS THAT
;  TITLE OF INVENTION:  INTERACT WITH AND/OR MODULATE PROTEIN TYROSINE KINASES AND/OR
;  TITLE OF INVENTION:  PROTEIN TYROSINE KINASE PATHWAYS IN BREAST CELLS
;  FILE REFERENCE: D0273A CIP
;  CURRENT APPLICATION NUMBER: US/11/072,175
;  CURRENT FILING DATE:  2005-03-05
;  PRIOR APPLICATION NUMBER: US 60/406,385
;  PRIOR FILING DATE: 2002-08-27
;  PRIOR APPLICATION NUMBER: US 10/648,593
;  PRIOR FILING DATE: 2003-08-26
;  NUMBER OF SEQ ID NOS: 571
;  SOFTWARE: PatentIn version 3.2
; SEQ ID NO 143
;    LENGTH: 339
;    TYPE: PRT
;    ORGANISM: Homo sapiens
US-11-072-175-143

  Query Match             11.9%;  Score 65.5;  DB 7;  Length 339;
  Best Local Similarity   43.2%;  Pred. No. 5.8;
  Matches   19;  Conservative   11;  Mismatches    9;  Indels    5;  Gaps    3;

Qy          9 LKKKAQYEANKVK--LWGLNTEYDNLLFVINS-IKQEI--VNRV 47
              ||  |||:|:::|  : || |: |:|: :| |   ||:  :|||
Db        103 LKTPAQYDASELKASMKGLGTDEDSLIEIICSRTNQELQEINRV 146


RESULT 12
US-11-207-626A-34
; Sequence 34, Application US/11207626A
; Publication No. US20060014276A1
; GENERAL INFORMATION:
;  APPLICANT: Havenga, Menzo
;  APPLICANT:  Vogels, Ronald
;  APPLICANT:  Bout, Abraham
;  TITLE OF INVENTION: CHIMERIC ADENOVIRUSES
;  FILE REFERENCE: 2578-4123.2US
;  CURRENT APPLICATION NUMBER: US/11/207,626A
;  CURRENT FILING DATE:  2005-08-18
;  PRIOR APPLICATION NUMBER: EP 98202297.2
;  PRIOR FILING DATE: 1998-07-08
;  NUMBER OF SEQ ID NOS: 87
;  SOFTWARE: PatentIn version 3.2
; SEQ ID NO 34
;    LENGTH: 378
;    TYPE: PRT
;    ORGANISM: Human Adenovirus 42 Fiber Protein
;    FEATURE:
;    NAME/KEY: misc_feature
;    LOCATION: (237)..(237)
;    OTHER INFORMATION: Xaa can be any amino acid
;    FEATURE:
;    NAME/KEY: misc_feature
;    LOCATION: (237)..(237)
;    OTHER INFORMATION: Xaa can be any naturally occurring amino acid
US-11-207-626A-34

  Query Match             11.9%;  Score 65.5;  DB 7;  Length 378;
  Best Local Similarity   25.3%;  Pred. No. 6.7;
  Matches   21;  Conservative   17;  Mismatches   34;  Indels   11;  Gaps    3;

Qy          1 KLAFRACRLKKKAQYEANKVKLWG-------LNTEYDNLL-FVINSIKQEIVNRVQNPRD 52
              || | |     | : |:|  : |         ::| ||| : |: |:   :|:|   :| :
Db        255 KLLFDA---NGKLKSESNLSQYWNYRSDNSVVSTPYDNAVPFMPNTAYPKIINSTTDPEN 311

Qy         53 ERGPNMGQKLEILIKDTLGLPVA 75
              ::        : ::   | |||
Db        312 KKSAKKTIVGNVYLEGNAGQPVA 334


RESULT 13
US-11-055-822-370
; Sequence 370, Application US/11055822
; Publication No. US20050260707A1
; GENERAL INFORMATION:
;  APPLICANT: Pompejus, Markus
;  APPLICANT:  Kroger, Burkhard
;  APPLICANT:  Schroder, Hartwig
;  APPLICANT:  Zelder, Oskar
;  APPLICANT:  Haberhauer, Gregor
;  TITLE OF INVENTION: CORYNEBACTERIUM GLUTAMICUM GENES ENCODING
;  TITLE OF INVENTION:  METABOLIC PATHWAY PROTEINS
;  FILE REFERENCE: BGI-121CPCN
;  CURRENT APPLICATION NUMBER: US/11/055,822
;  CURRENT FILING DATE:  2005-02-11
;  PRIOR APPLICATION NUMBER: 09/606,740
;  PRIOR FILING DATE: 2000-06-23
;  PRIOR APPLICATION NUMBER: 60/141,031
;  PRIOR FILING DATE: 1999-06-25
;  PRIOR APPLICATION NUMBER: 60/142,101
;  PRIOR FILING DATE: 1999-07-02
;  PRIOR APPLICATION NUMBER: 60/148,613
;  PRIOR FILING DATE: 1999-08-12
;  PRIOR APPLICATION NUMBER: 60/187,970
;  PRIOR FILING DATE: 2000-03-09
;  PRIOR APPLICATION NUMBER: DE 19930476.9
;  PRIOR FILING DATE: 1999-07-01
;  PRIOR APPLICATION NUMBER: DE 19931415.2
;  PRIOR FILING DATE: 1999-07-08
;  PRIOR APPLICATION NUMBER: DE 19931418.7
;  PRIOR FILING DATE: 1999-07-08
;  PRIOR APPLICATION NUMBER: DE 19931419.5
;  PRIOR FILING DATE: 1999-07-08
;  PRIOR APPLICATION NUMBER: DE 19931420.9
;  PRIOR FILING DATE: 1999-07-08
;  Remaining Prior Application data removed - See File Wrapper or PALM.
;  NUMBER OF SEQ ID NOS: 1158
; SEQ ID NO 370
;    LENGTH: 1066
;    TYPE: PRT
;    ORGANISM: Corynebacterium glutamicum
US-11-055-822-370

  Query Match             11.9%;  Score 65.5;  DB 7;  Length 1066;
  Best Local Similarity   27.5%;  Pred. No. 26;
  Matches   25;  Conservative   11;  Mismatches   38;  Indels   17;  Gaps    4;

Qy         24 GLNTEYDNLLFVINSIKQEIVNRVQNPRDERGPNMGQKLEILIKDTLGLPVA-------- 75
              |::|    |: | || :    |:     ||| |   :  |          | |:|
Db        234 GVDTGGCNIQFAINPVDGRIITIEMNPRVSRSSALASK-------ATGFPIAKMAAKLAI 286
```

```
Qy          76 GQT-SEFVNQVLEKT-AEGNPTGGLVGLRIP 104
               |  |  |  |  :  :| |   ||   | :: |
Db         287 GYTLDEITNDITGETPAAFEPTIDYVVVKAP 317


RESULT 14
US-11-055-822-1002
; Sequence 1002, Application US/11055822
; Publication No. US20050260707A1
; GENERAL INFORMATION:
;  APPLICANT: Pompejus, Markus
;  APPLICANT:  Kroger, Burkhard
;  APPLICANT:  Schroder, Hartwig
;  APPLICANT:  Zelder, Oskar
;  APPLICANT:  Haberhauer, Gregor
;  TITLE OF INVENTION: CORYNEBACTERIUM GLUTAMICUM GENES ENCODING
;  TITLE OF INVENTION:  METABOLIC PATHWAY PROTEINS
;  FILE REFERENCE: BGI-121CPCN
;  CURRENT APPLICATION NUMBER: US/11/055,822
;  CURRENT FILING DATE:  2005-02-11
;  PRIOR APPLICATION NUMBER: 09/606,740
;  PRIOR FILING DATE: 2000-06-23
;  PRIOR APPLICATION NUMBER: 60/141,031
;  PRIOR FILING DATE: 1999-06-25
;  PRIOR APPLICATION NUMBER: 60/142,101
;  PRIOR FILING DATE: 1999-07-02
;  PRIOR APPLICATION NUMBER: 60/148,613
;  PRIOR FILING DATE: 1999-08-12
;  PRIOR APPLICATION NUMBER: 60/187,970
;  PRIOR FILING DATE: 2000-03-09
;  PRIOR APPLICATION NUMBER: DE 19930476.9
;  PRIOR FILING DATE: 1999-07-01
;  PRIOR APPLICATION NUMBER: DE 19931415.2
;  PRIOR FILING DATE: 1999-07-08
;  PRIOR APPLICATION NUMBER: DE 19931418.7
;  PRIOR FILING DATE: 1999-07-08
;  PRIOR APPLICATION NUMBER: DE 19931419.5
;  PRIOR FILING DATE: 1999-07-08
;  PRIOR APPLICATION NUMBER: DE 19931420.9
;  PRIOR FILING DATE: 1999-07-08
;  Remaining Prior Application data removed - See File Wrapper or PALM.
;  NUMBER OF SEQ ID NOS: 1158
; SEQ ID NO 1002
;   LENGTH: 1066
;   TYPE: PRT
;   ORGANISM: Corynebacterium glutamicum
US-11-055-822-1002

  Query Match             11.9%;  Score 65.5;  DB 7;  Length 1066;
  Best Local Similarity   27.5%;  Pred. No. 26;
  Matches   25;  Conservative   11;  Mismatches   38;  Indels   17;  Gaps    4;

Qy          24 GLNTEYDNLLFVINSIKQEIVNRVQNPRDERGPNMGQKLEILIKDTLGLPVA-------- 75
               |::|   |: | || :   |:    |||  |   :  |         | |:|
Db         234 GVDTGGCNIQFAINPVDGRIITIEMNPRVSRSSALASK-------ATGFPIAKMAAKLAI 286

Qy          76 GQT-SEFVNQVLEKT-AEGNPTGGLVGLRIP 104
               |  |  |  |  :  :| |   ||   | :: |
Db         287 GYTLDEITNDITGETPAAFEPTIDYVVVKAP 317


RESULT 15
US-11-055-822-368
; Sequence 368, Application US/11055822
; Publication No. US20050260707A1
; GENERAL INFORMATION:
;  APPLICANT: Pompejus, Markus
;  APPLICANT:  Kroger, Burkhard
;  APPLICANT:  Schroder, Hartwig
;  APPLICANT:  Zelder, Oskar
;  APPLICANT:  Haberhauer, Gregor
;  TITLE OF INVENTION: CORYNEBACTERIUM GLUTAMICUM GENES ENCODING
;  TITLE OF INVENTION:  METABOLIC PATHWAY PROTEINS
;  FILE REFERENCE: BGI-121CPCN
;  CURRENT APPLICATION NUMBER: US/11/055,822
;  CURRENT FILING DATE:  2005-02-11
;  PRIOR APPLICATION NUMBER: 09/606,740
;  PRIOR FILING DATE: 2000-06-23
;  PRIOR APPLICATION NUMBER: 60/141,031
;  PRIOR FILING DATE: 1999-06-25
;  PRIOR APPLICATION NUMBER: 60/142,101
;  PRIOR FILING DATE: 1999-07-02
;  PRIOR APPLICATION NUMBER: 60/148,613
;  PRIOR FILING DATE: 1999-08-12
;  PRIOR APPLICATION NUMBER: 60/187,970
;  PRIOR FILING DATE: 2000-03-09
;  PRIOR APPLICATION NUMBER: DE 19930476.9
;  PRIOR FILING DATE: 1999-07-01
;  PRIOR APPLICATION NUMBER: DE 19931415.2
;  PRIOR FILING DATE: 1999-07-08
;  PRIOR APPLICATION NUMBER: DE 19931418.7
;  PRIOR FILING DATE: 1999-07-08
;  PRIOR APPLICATION NUMBER: DE 19931419.5
;  PRIOR FILING DATE: 1999-07-08
;  PRIOR APPLICATION NUMBER: DE 19931420.9
;  PRIOR FILING DATE: 1999-07-08
;  Remaining Prior Application data removed - See File Wrapper or PALM.
;  NUMBER OF SEQ ID NOS: 1158
; SEQ ID NO 368
;   LENGTH: 1113
;   TYPE: PRT
;   ORGANISM: Corynebacterium glutamicum
US-11-055-822-368

  Query Match             11.9%;  Score 65.5;  DB 7;  Length 1113;
  Best Local Similarity   27.5%;  Pred. No. 28;
  Matches   25;  Conservative   11;  Mismatches   38;  Indels   17;  Gaps    4;

Qy          24 GLNTEYDNLLFVINSIKQEIVNRVQNPRDERGPNMGQKLEILIKDTLGLPVA-------- 75
               |::|   |: | || :   |:    |||  |   :  |         | |:|
Db         281 GVDTGGCNIQFAINPVDGRIITIEMNPRVSRSSALASK-------ATGFPIAKMAAKLAI 333

Qy          76 GQT-SEFVNQVLEKT-AEGNPTGGLVGLRIP 104
               |  |  |  |  :  :| |   ||   | :: |
Db         334 GYTLDEITNDITGETPAAFEPTIDYVVVKAP 364


Search completed: February 28, 2006, 09:02:41
Job time : 3.88722 secs
```

This Page Blank (uspto)

GenCore version 5.1.7

OM protein - protein search, using sw model

Run on: February 28, 2006, 08:45:35 ; Search time 6.58647 Seconds
(without alignments)
1577.691 Million cell updates/sec

Title: US-10-717-665A-44_COPY_532_639
Perfect score: 551
Sequence: 1 KLAFRACRLKKKAQYEANKV.....TAEGNPTGGLVGLRIPTSKV 108

Scoring table: BLOSUM62
Gapop 10.0 , Gapext 0.5

Searched: 283416 seqs, 96216763 residues

Total number of hits satisfying chosen parameters: 283416

Minimum DB seq length: 0
Maximum DB seq length: 2000000000

Post-processing: Minimum Match 0%
Maximum Match 100%
Listing first 45 summaries

Database : PIR_80:*
1: pir1:*
2: pir2:*
3: pir3:*
4: pir4:*

Pred. No. is the number of results predicted by chance to have a score greater than or equal to the score of the result being printed, and is derived by analysis of the total score distribution.

SUMMARIES

Result No.	Score	% Query Match	Length	DB	ID	Description
1	82	14.9	571	2	AH2067	two-component hybr
2	82	14.9	902	2	T00072	hypothetical prote
3	79.5	14.4	534	2	F96713	unknown protein T6
4	78.5	14.2	281	2	E70173	dimethyladenosine
5	76.5	13.9	455	2	S65157	hypothetical prote
6	75.5	13.7	642	2	G95129	1,4-alpha-glucan b
7	74	13.4	869	2	A86983	conserved hypothet
8	74	13.4	869	2	S72760	pps1 protein - Myc
9	73.5	13.3	1104	2	AB2282	carbamoyl phosphat
10	73.5	13.3	1495	2	S60255	transcription co-r
11	72.5	13.2	509	2	T08436	inositol-3-phospha
12	72.5	13.2	642	2	E98000	1,4-alpha-glucan b
13	72	13.1	718	2	T49572	related to SHK1 KI
14	71	12.9	505	2	JN0760	H+-transporting tw
15	71	12.9	1355	2	T22552	hypothetical prote
16	70.5	12.8	667	2	S48285	probable glycine-t
17	70	12.7	200	2	AD2116	nicotinate-nucleot
18	70	12.7	846	2	H70871	hypothetical prote
19	69.5	12.6	1105	2	S76557	carbamoyl-phosphat
20	69	12.5	315	2	AH2712	conserved hypothet
21	69	12.5	315	2	F97494	hypothetical prote
22	69	12.5	330	2	F82338	hypothetical prote
23	69	12.5	803	2	A35954	endoplasmin precur
24	68.5	12.4	222	2	C34223	transcription fact
25	68.5	12.4	386	2	E75588	glutaryl-CoA dehyd
26	68.5	12.4	552	2	T23171	hypothetical prote
27	68	12.3	258	2	S71561	drought-induced pr
28	68	12.3	1328	2	AE2351	protoporphyrin IX
29	67.5	12.3	186	2	F64323	ribosomal protein
30	67.5	12.3	339	2	S33700	annexin II - rat
31	67.5	12.3	544	2	F82366	probable ubiquinon
32	67.5	12.3	802	2	A29317	endoplasmic reticu
33	67.5	12.3	804	2	A53211	glucose-regulated
34	67.5	12.3	804	2	S51358	protein kinase ppk
35	67.5	12.3	858	2	A57513	heat shock protein
36	67.5	12.3	1340	2	A39808	proteoglycan core
37	67.5	12.3	2327	2	T42630	aggrecan - bovine
38	67	12.2	588	2	A82232	malate oxidoreduct
39	67	12.2	878	1	RRXSIB	RNA-directed RNA p
40	67	12.2	1130	2	T34081	hypothetical prote
41	67	12.2	13055	2	T16580	hypothetical prote
42	66.5	12.1	181	2	A54025	transcription fact
43	66.5	12.1	181	2	A39382	liver regeneration
44	66.5	12.1	339	1	LUBO36	annexin II - bovin
45	66	12.0	224	2	B86850	transcription regu

ALIGNMENTS

RESULT 1
AH2067
two-component hybrid sensor and regulator all2094 [imported] - Nostoc sp. (strain PCC 71
C;Species: Nostoc sp. PCC 7120
A;Note: Nostoc sp. strain PCC 7120 is a synonym of Anabaena sp. strain PCC 7120
C;Date: 14-Dec-2001 #sequence_revision 14-Dec-2001 #text_change 09-Jul-2004
C;Accession: AH2067
R;Kaneko, T.; Nakamura, Y.; Wolk, C.P.; Kuritz, T.; Sasamoto, S.; Watanabe, A.; Iriguchi, Nakazaki, N.; Shimpo, S.; Sugimoto, M.; Takazawa, M.; Yamada, M.; Yasuda, M.; Tabata, S.
DNA Res. 8, 205-213, 2001
A;Title: Complete Genomic Sequence of the Filamentous Nitrogen-fixing Cyanobacterium Anat
A;Reference number: AB1807; MUID:21595285; PMID:11759840
A;Accession: AH2067
A;Status: preliminary
A;Molecule type: DNA
A;Residues: 1-571 <KUR>
A;Cross-references: UNIPROT:Q8YV87; UNIPARC:UPI00000CE2FB; GB:BA000019; PIDN:BAB73793.1;
A;Experimental source: strain PCC 7120
C;Genetics:
A;Gene: all2094

Query Match 14.9%; Score 82; DB 2; Length 571;
Best Local Similarity 24.2%; Pred. No. 2.4;
Matches 24; Conservative 17; Mismatches 38; Indels 20; Gaps 3;

```
Qy          23 WGLNTEYD-NLLFVINSIKQEIVNRVQNP------------RDERGPNMGQKLEILIKD 68
               |  :   | ||  |:  | :: :| :|||             |   | :| :: :: ||
Db         172 WQFDDAQDRNLTEVVQLIDEQTLNSIQNPIISAIQKETTVYLGDGLSPTVGHRIMLITKD 231

Qy          69 TLGLPVAGQTSEFVNQVLEKTAEGNPTGGLVGLRIPTSK 107
                  :|||   :   |          |: || ::  |  | :
Db         232 GTMIPVADSATPLRNN------NGDITGAVMVFRDDTQR 264
```

RESULT 2
T00072
hypothetical protein KIAA0528 - human
C;Species: Homo sapiens (man)
C;Date: 22-Jan-1999 #sequence_revision 22-Jan-1999 #text_change 31-Dec-2004
C;Accession: T00072
R;Nagase, T.; Ishikawa, K.; Miyajima, N.; Tanaka, A.; Kotani, H.; Nomura, N.; Ohara, O.
DNA Res. 5, 31-39, 1998
A;Title: Prediction of the coding sequences of unidentified human genes. IX. The complete
A;Reference number: Z14086; MUID:98290545; PMID:9628581
A;Accession: T00072
A;Status: preliminary; translated from GB/EMBL/DDBJ
A;Molecule type: mRNA
A;Residues: 1-902 <NAG>
A;Cross-references: UNIPROT:O60280; UNIPARC:UPI0000178FFB; EMBL:AB011100; PIDN:BAA25454.1
A;Experimental source: brain; clone HG2576
C;Genetics:
A;Note: KIAA0528

```
  Query Match             14.9%;  Score 82;  DB 2;  Length 902;
  Best Local Similarity   28.4%;  Pred. No. 4.3;
  Matches   29;  Conservative   22;  Mismatches   27;  Indels   24;  Gaps    6;

Qy          5 RACRLKKKAQYEANKVKLWGLNTEYDNLL-FVINSIKQEIVNRVQNPRDERGPN--MGQK 61
              | |||||||| |||          |   ||| |:   :  :::|:::     :| |   | :
Db        368 RLCRLKKKAQAEANA-------TAISNLLPFMEYEVHTQLMNKLK----LKGMNALFGLR 416

Qy         62 LEILIKDTLGLPVAGQTSEFVNQVLEKTAEGNPTGGLVGLRI 103
              ::| : : : : :|  |  ::  :        || |  |::|
Db        417 IQITVGENMLMGLASATGVYLAAL--------PTPG--GIQI 448


RESULT 3
F96713
unknown protein T6L1.9 [imported] - Arabidopsis thaliana
C;Species: Arabidopsis thaliana (mouse-ear cress)
C;Date: 02-Mar-2001 #sequence_revision 02-Mar-2001 #text_change 09-Jul-2004
C;Accession: F96713
R;Theologis, A.; Ecker, J.R.; Palm, C.J.; Federspiel, N.A.; Kaul, S.; White, O.; Alonso,
 Chin, C.W.; Chung, M.K.; Conn, L.; Conway, A.B.; Conway, A.R.; Creasy, T.H.; Dewar, K.;
ansen, N.F.; Hughes, B.; Huizar, L.
Nature 408, 816-820, 2000
A;Authors: Hunter, J.L.; Jenkins, J.; Johnson-Hopson, C.; Khan, S.; Khaykin, E.; Kim, C.
C.A.; Li, J.H.; Li, Y.; Lin, X.; Liu, S.X.; Liu, Z.A.; Luros, J.S.; Maiti, R.; Marziali,
Rizzo, M.; Rooney, T.; Rowley, D.; Sakano, H.
A;Authors: Salzberg, S.L.; Schwartz, J.R.; Shinn, P.; Southwick, A.M.; Sun, H.; Tallon,
ker, M.; Wu, D.; Yu, G.; Fraser, C.M.; Venter, J.C.; Davis, R.W.
A;Title: Sequence and analysis of chromosome 1 of the plant Arabidopsis.
A;Reference number: A86141; MUID:21016719; PMID:11130712
A;Accession: F96713
A;Status: preliminary
A;Molecule type: DNA
A;Residues: 1-534 <STO>
A;Cross-references: UNIPROT:Q9CAB0; UNIPARC:UPI00000A0D17; GB:AE005173; NID:g6358763; PI
C;Genetics:
A;Gene: T6L1.9
A;Map position: 1

  Query Match             14.4%;  Score 79.5;  DB 2;  Length 534;
  Best Local Similarity   30.4%;  Pred. No. 4;
  Matches   34;  Conservative   16;  Mismatches   33;  Indels   29;  Gaps    6;

Qy          8 RLKKKAQY---------EANKVKLWGLNTEYDNLLFVINSIKQEIVNRVQ------NPRD 52
              :|| | |           ||::  :||   | ||   |:  | :|:| |:|      |
Db        211 QLKLKLHYTEEVSSRMEEASEF-IWGRFLEADNSSEVLTGISKELVGRLQILQFSLNGSA 269

Qy         53 ERGPNMGQKLE-----ILIKDTLGLPVAG---QTSEFVNQVLE-----KTAE 91
              :|   :  |||     :  || |   : |   : || |::||       |:||
Db        270 QRESELKSKLEDCTVQLEAKDLLVQKLEGTISENSEIVSEVLTLREYVKSAE 321


RESULT 4
E70173
dimethyladenosine transferase (ksgA) homolog - Lyme disease spirochete
C;Species: Borrelia burgdorferi (Lyme disease spirochete)
C;Date: 13-Feb-1998 #sequence_revision 13-Feb-1998 #text_change 31-Dec-2004
C;Accession: E70173
R;Fraser, C.M.; Casjens, S.; Huang, W.M.; Sutton, G.G.; Clayton, R.; Lathigra, R.; White
son, D.; Peterson, J.; Kerlavage, A.R.; Quackenbush, J.; Salzberg, S.; Hanson, M.; Vugt,
; Bowman, C.; Garland, S.; Fujii, C.; Cotton, M.D.; Horst, K.; Roberts, K.; Hatch, B.
Nature 390, 580-586, 1997
A;Authors: Smith, H.O.; Venter, J.C.
A;Title: Genomic sequence of a Lyme disease spirochaete, Borrelia burgdorferi.
A;Reference number: A70100; MUID:98065943; PMID:9403685
A;Accession: E70173
A;Status: preliminary; nucleic acid sequence not shown; translation not shown
A;Molecule type: DNA
A;Residues: 1-281 <KLE>
A;Cross-references: UNIPROT:O51536; UNIPARC:UPI0000057535; GB:AE001160; GB:AE000783; NID
A;Experimental source: strain B31
C;Superfamily: dimethyladenosine transferase (rRNA adenosine dimethyltransferase)

  Query Match             14.2%;  Score 78.5;  DB 2;  Length 281;
  Best Local Similarity   31.6%;  Pred. No. 2.3;
  Matches   24;  Conservative   14;  Mismatches   23;  Indels   15;  Gaps    4;

Qy         21 KLWGLNTEYDNLLFVIN-SIKQEIVNRVQNPRDER----GPNMGQKLEILIKDT---LGL 72
              |||| |       ::|| ||:|:|:  :    :|:    || :|   |||:| |
Db         31 KLWGQN-------YLINESIRQKIIESLDIKENEKIWEIGPGLGAMTEILLKKTNLLTAF 83

Qy         73 PVAGQTSEFVNQVLEK 88
              :  : || :|:   |
Db         84 EIDLKYSEILNEKFGK 99


RESULT 5
S65157
hypothetical protein YPL146c - yeast (Saccharomyces cerevisiae)
N;Alternate names: hypothetical protein P2610
C;Species: Saccharomyces cerevisiae
C;Date: 10-Dec-1994 #sequence_revision 31-May-1996 #text_change 31-Dec-2004
C;Accession: S65157; S69041; S69453
R;Purnelle, B.; Coster, F.; Goffeau, A.
submitted to the Protein Sequence Database, May 1996
A;Reference number: S65154
A;Accession: S65157
A;Molecule type: DNA
A;Residues: 1-455 <PUR>
A;Cross-references: UNIPROT:Q12080; UNIPARC:UPI000013BD31; EMBL:Z73502; NID:g1370311; PII
A;Experimental source: strain S288C (AB972)
R;Hall, J.; DePaulo, T.; Ahmed, A.; Bussey, H.; Fortin, N.; Friesen, J.D.; Storms, R.K.;
submitted to the EMBL Data Library, December 1995
A;Description: The sequence of Saccharomyces cerevisiae chromosome XVI left arm.
A;Reference number: S69040
A;Accession: S69041
A;Molecule type: DNA
A;Residues: 1-455 <HAL>
A;Cross-references: UNIPARC:UPI000013BD31; EMBL:U43703; NID:g1244769; PID:g1244771; MIPS:
R;Purnelle, B.; Comblez, S.; Coster, F.; Naveau, F.; Goffeau, A.
submitted to the EMBL Data Library, March 1996
A;Description: The sequence of 55 kb on the left arm of yeast chromosome XVI identifies 7
ogue to the human phosphotyrosyl phosphatase activator PTPA and a homologue to the plant
A;Reference number: S69428
A;Accession: S69453
A;Molecule type: DNA
A;Residues: 1-455 <PUW>
A;Cross-references: UNIPARC:UPI000013BD31; EMBL:X96770; NID:g1403537; PID:g1403563
C;Genetics:
A;Cross-references: SGD:S0006067
A;Map position: 16L
C;Superfamily: tumor suppressor protein, Gltscr2/p60 type

  Query Match             13.9%;  Score 76.5;  DB 2;  Length 455;
  Best Local Similarity   24.8%;  Pred. No. 6.7;
  Matches   31;  Conservative   21;  Mismatches   42;  Indels   31;  Gaps    5;

Qy         10 KKKAQYEANKVK----------LWGLNTEYDNLLFVINSIKQEIV-------NRVQNPR 51
              ||| :|: || |           |  |      :|  |||| : ||:       |:|:  :
Db        317 KKKTKYQRNKAKRHEEKVKLQQELKELRQRVKDLEEVINSEETEILSAIESDSNKVKKSK 376

Qy         52 DERGPNMGQKLEILIKDTLGLPVAGQTSEFVNQVLEKTAEGN---------PTGGLVGLR 102
              :   :| |  : | : | :  : : |: : ::     |||          : | | |
Db        377 KNKKHKLGTKYSV-IDERLEIKFSDELSDSLRKL---KPEGNLLYDTVRKLQSSGKVETR 432

Qy        103 IPTSK 107
              :| |
Db        433 VPVRK 437


RESULT 6
G95129
1,4-alpha-glucan branching enzyme [imported] - Streptococcus pneumoniae (strain TIGR4)
C;Species: Streptococcus pneumoniae
C;Date: 03-Aug-2001 #sequence_revision 03-Aug-2001 #text_change 09-Jul-2004
```

```
C;Accession: G95129
R;Tettelin, H.; Nelson, K.E.; Paulsen, I.T.; Eisen, J.A.; Read, T.D.; Peterson, S.; Heid
on, J.D.; Umayam, L.A.; White, O.; Salzberg, S.L.; Lewis, M.R.; Radune, D.; Holtzapple,
nson, T.; Hickey, E.K.; Holt, I.E.
Science 293, 498-506, 2001
A;Authors: Loftus, B.J.; Yang, F.; Smith, H.O.; Venter, J.C.; Dougherty, B.A.; Morrison,
A;Title: Complete Genome Sequence of a virulent isolate of Streptococcus pneumoniae.
A;Reference number: A95000; MUID:21357209; PMID:11463916
A;Accession: G95129
A;Status: preliminary
A;Molecule type: DNA
A;Residues: 1-642 <KUR>
A;Cross-references: UNIPROT:Q97QS8; UNIPARC:UPI00000C9CBF; GB:AE005672; PIDN:AAK75232.1;
A;Experimental source: strain TIGR4
C;Genetics:
A;Gene: SP1121
C;Superfamily: 1,4-alpha-glucan branching enzyme

  Query Match             13.7%;  Score 75.5;  DB 2;  Length 642;
  Best Local Similarity   27.6%;  Pred. No. 13;
  Matches   29;  Conservative   17;  Mismatches   20;  Indels   39;  Gaps    6;

Qy          9 LKKKAQYEANKVK--------LWGLNTEYDNLLFVINSIKQEIVNRVQNPRDERGPNM-- 58
              :  | :| |:::          || ::| || :        ||::      | |  ::
Db        490 MNAKMKYFASQLNQFYKDHRCLWEIDTSYDGI---------EIID-----ADNRDQSVLS 535

Qy         59 ----GQKLEILI----------KD-TLGLPVAGQTSEFVNQVLEK 88
                  |:| |:|:          || |:||||||   |  |  ||:
Db        536 FIRKGKKGEMLVCIFNMVPVERKDFTIGLPVAGIYEEVWNTELEE 580


RESULT 7
A86983
conserved hypothetical protein ML0593 [imported] - Mycobacterium leprae
C;Species: Mycobacterium leprae
C;Date: 20-Apr-2001 #sequence_revision 20-Apr-2001 #text_change 09-Jul-2004
C;Accession: A86983
R;Cole, S.T.; Eiglmeier, K.; Parkhill, J.; James, K.D.; Thomson, N.R.; Wheeler, P.R.; Ho
R.; Davies, R.M.; Devlin, K.; Duthoy, S.; Feltwell, T.; Fraser, A.; Hamlin, N.; Holroyd,
eam, M.A.; Rutherford, K.M.
Nature 409, 1007-1011, 2001
A;Authors: Rutter, S.; Seeger, K.; Simon, S.; Simmonds, M.; Skelton, J.; Squares, R.; Sq
A;Title: Massive gene decay in the leprosy bacillus.
A;Reference number: A86909; MUID:21128732; PMID:11234002
A;Accession: A86983
A;Status: preliminary
A;Molecule type: DNA
A;Residues: 1-869 <STO>
A;Cross-references: UNIPROT:Q49689; UNIPARC:UPI000013AB0A; GB:AL450380; NID:g13092788; P
C;Genetics:
A;Gene: ML0593

  Query Match             13.4%;  Score 74;  DB 2;  Length 869;
  Best Local Similarity   23.6%;  Pred. No. 26;
  Matches   21;  Conservative   19;  Mismatches   29;  Indels   20;  Gaps    3;

Qy          8 RLKKKAQYEANKVKLWGLN---TEYDNLLFVINSIKQEIVNRVQNPRDERGPNMGQKLEI 64
              |||    :|   :  || |    ::||: : : | :::  :  : | | |
Db         68 RLKALRVFERKPMPRWGSNLDGIDFDNIKYFVRSTEKQAASWDELPEDIRN--------- 118

Qy         65 LIKDTLGLP-------VAGQTSEFVNQVL 86
                | ||:|       |||  ::: ::|:
Db        119 -TYDRLGIPDAEKQRLVAGVAAQYESEVV 146


RESULT 8
S72760
pps1 protein - Mycobacterium leprae
N;Alternate names: B1496_C2_189 protein
C;Species: Mycobacterium leprae
C;Date: 19-Mar-1997 #sequence_revision 25-Apr-1997 #text_change 09-Jul-2004
C;Accession: S72760; T11013
R;Smith, D.R.; Robison, K.
submitted to the EMBL Data Library, November 1993
A;Description: Mycobacterium leprae cosmid B1496.
A;Reference number: S72695
A;Accession: S72760
A;Status: preliminary
A;Molecule type: DNA
A;Residues: 1-869 <SMI>
A;Cross-references: UNIPROT:Q49689; UNIPARC:UPI000016FA8B; EMBL:U00013; NID:g466868; PIDN
R;Parkhill, J.; Barrell, B.G.; Rajandream, M.A.
submitted to the EMBL Data Library, September 1997
A;Reference number: Z16918
A;Accession: T11013
A;Status: preliminary; translated from GB/EMBL/DDBJ
A;Molecule type: DNA
A;Residues: 202-481,'A',483-589 <PAR>
A;Cross-references: UNIPARC:UPI000016FB03; EMBL:Z99125; NID:g2398683; PIDN:CAB16172.1; PI

  Query Match             13.4%;  Score 74;  DB 2;  Length 869;
  Best Local Similarity   23.6%;  Pred. No. 26;
  Matches   21;  Conservative   19;  Mismatches   29;  Indels   20;  Gaps    3;

Qy          8 RLKKKAQYEANKVKLWGLN--TEYDNLLFVINSIKQEIVNRVQNPRDERGPNMGQKLEI 64
              |||    :|   :  || |    ::||: : : | :::  :  : | | |
Db         68 RLKALRVFERKPMPRWGSNLDGIDFDNIKYFVRSTEKQAASWDELPEDIRN--------- 118

Qy         65 LIKDTLGLP-------VAGQTSEFVNQVL 86
                | ||:|       |||  ::: ::|:
Db        119 -TYDRLGIPDAEKQRLVAGVAAQYESEVV 146


RESULT 9
AB2282
carbamoyl phosphate synthetase large chain [imported] - Nostoc sp. (strain PCC 7120)
C;Species: Nostoc sp. PCC 7120
A;Note: Nostoc sp. strain PCC 7120 is a synonym of Anabaena sp. strain PCC 7120
C;Date: 14-Dec-2001 #sequence_revision 14-Dec-2001 #text_change 09-Jul-2004
C;Accession: AB2282
R;Kaneko, T.; Nakamura, Y.; Wolk, C.P.; Kuritz, T.; Sasamoto, S.; Watanabe, A.; Iriguchi,
 Nakazaki, N.; Shimpo, S.; Sugimoto, M.; Takazawa, M.; Yamada, M.; Yasuda, M.; Tabata, S.
DNA Res. 8, 205-213, 2001
A;Title: Complete Genomic Sequence of the Filamentous Nitrogen-fixing Cyanobacterium Anal
A;Reference number: AB1807; MUID:21595285; PMID:11759840
A;Accession: AB2282
A;Status: preliminary
A;Molecule type: DNA
A;Residues: 1-1104 <KUR>
A;Cross-references: UNIPROT:Q8YQL2; UNIPARC:UPI0000126F6F; GB:BA000019; PIDN:BAB75508.1;
A;Experimental source: strain PCC 7120
C;Genetics:
A;Gene: alr3809
C;Superfamily: carbamoyl-phosphate synthase (glutamine-hydrolyzing) large chain; biotin c

  Query Match             13.3%;  Score 73.5;  DB 2;  Length 1104;
  Best Local Similarity   27.5%;  Pred. No. 39;
  Matches   25;  Conservative   12;  Mismatches   37;  Indels   17;  Gaps    4;

Qy         24 GLNTEYDNLLFVINSIKQEIVNRVQNPRDERGPNMGQKLEILIKDTLGLPVA-------- 75
              |: |   |: | :| :  ::|    |||  |   :  |          | |:|
Db        277 GVETGGSNIQFAVNPVNGDVVVIEMNPRVSRSSALSSK-------ATGFPIAKMAAKLAV 329

Qy         76 GQT-SEFVNQVLEKT-AEGNPTGGLVGLRIP 104
              | |  |  | : :|| |   ||   |  ::|
Db        330 GYTLDEIRNDITKKTPASFEPTIDYVVTKVP 360


RESULT 10
S60255
transcription co-repressor SMRT - human
C;Species: Homo sapiens (man)
C;Date: 10-Apr-1996 #sequence_revision 19-Apr-1996 #text_change 09-Jul-2004
C;Accession: S60255
R;Chen, J.D.; Evans, R.M.
Nature 377, 454-457, 1995
```

```
A;Title: A transcriptional co-repressor that interacts with nuclear hormone receptors.
A;Reference number: S60255; MUID:96008552; PMID:7566127
A;Accession: S60255
A;Status: preliminary; nucleic acid sequence not shown
A;Molecule type: mRNA
A;Residues: 1-1495 <CHE>
A;Cross-references: UNIPROT:Q9Y618; UNIPARC:UPI000016A25F; EMBL:U37146; NID:g1045654; PI

  Query Match             13.3%;  Score 73.5;  DB 2;  Length 1495;
  Best Local Similarity   30.6%;  Pred. No. 57;
  Matches   22;  Conservative   14;  Mismatches   17;  Indels   19;  Gaps    3;

Qy         37 NSIKQEIVNRVQNPRDERGPNMGQKLEILIKDTLGLPVAGQTSEFVNQVLEKTAEGNPTG 96
              : :||| :    :|| : ||           ::||:| | : |     ||  || |:  |
Db        143 SGVKQEQL----SPRGQAGP----------PESLGVPTAQEAS-----VLRGTALGSVPG 183

Qy         97 GLVGLRIPTSKV 108
              | :   ||:::|
Db        184 GSITKGIPSTRV 195


RESULT 11
T08436
inositol-3-phosphate synthase (EC 5.5.1.4) [similarity] - rape
C;Species: Brassica napus (rape)
C;Date: 11-Jun-1999 #sequence_revision 11-Jun-1999 #text_change 09-Jul-2004
C;Accession: T08436
R;Hussain, A.; Bourgeois, J.; Polvi, S.; Tsang, E.; Keller, W.A.; Georges, F.
submitted to the EMBL Data Library, August 1996
A;Reference number: Z16418
A;Accession: T08436
A;Status: preliminary; translated from GB/EMBL/DDBJ
A;Molecule type: mRNA
A;Residues: 1-509 <HUS>
A;Cross-references: UNIPROT:Q96348; UNIPARC:UPI000017952A; EMBL:U66307; NID:g1513227; PI
C;Function:
A;Description: catalyzes the reversible isomerization of D-glucose 6-phosphate to 1L-myc
C;Superfamily: myo-inositol-1-phosphate synthase
C;Keywords: intramolecular lyase; isomerase

  Query Match             13.2%;  Score 72.5;  DB 2;  Length 509;
  Best Local Similarity   26.8%;  Pred. No. 19;
  Matches   26;  Conservative   15;  Mismatches   29;  Indels   27;  Gaps    6;

Qy         16 EANKVK----LWGLNTE-YDNLLFVINSIKQEIVNRVQNPRDERGPNMGQKLEILIKDTL 70
              | |||     ||  ||| | |::  :|   : ::| |      |  |:    :  :::
Db        217 EKNKVDKVVVLWTANTERYSNVIVGLNDTMENLMNSVDRDESEISPSTLYAIACVLE--- 273

Qy         71 GLPVAGQTSEFVNQVLEKTAEGNPTG----GLVGLRI 103
              |:|       |:|        |:|      ||: | |
Db        274 GIP-------FIN--------GSPQNTFVPGLIDLAI 295


RESULT 12
E98000
1,4-alpha-glucan branching enzyme (EC 2.4.1.18) [imported] - Streptococcus pneumoniae (s
C;Species: Streptococcus pneumoniae
C;Date: 22-Oct-2001 #sequence_revision 22-Oct-2001 #text_change 09-Jul-2004
C;Accession: E98000
R;Hoskins, J.A.; Alborn Jr., W.; Arnold, J.; Blaszczak, L.; Burgett, S.; DeHoff, B.S.; E
e, R.; LeBlanc, D.J.; Lee, L.N.; Lefkowitz, E.J.; Lu, J.; Matsushima, P.; McAhren, S.; M
y, P.; Sun, P.M.; Winkler, M.E.
J. Bacteriol. 183, 5709-5717, 2001
A;Authors: Yang, Y.; Young-Bellido, M.; Zhao, G.; Zook, C.; Baltz, R.H.; Jaskunas, S.R.;
A;Title: Genome of the Bacterium Streptococcus pneumoniae Strain R6.
A;Reference number: A97872; MUID:21429245; PMID:11544234
A;Accession: E98000
A;Status: preliminary
A;Molecule type: DNA
A;Residues: 1-642 <KUR>
A;Cross-references: UNIPROT:Q8DPS6; UNIPARC:UPI00000E3597; GB:AE007317; PIDN:AAK99833.1;
C;Genetics:
A;Gene: glgB
```

```
C;Superfamily: 1,4-alpha-glucan branching enzyme
C;Keywords: glycosyltransferase; hexosyltransferase

  Query Match             13.2%;  Score 72.5;  DB 2;  Length 642;
  Best Local Similarity   31.0%;  Pred. No. 26;
  Matches   26;  Conservative   12;  Mismatches   15;  Indels   31;  Gaps    5;

Qy         22 LWGLNTEYDNLLFVINSIKQEIVNRVQNPRDERGPNM------GQKLEILI--------- 66
              || ::| || :          ||::      | |  ::      |:| |:|:
Db        511 LWEIDTSYDGI---------EIID-----ADNRDQSVLSFIRKGKKGEMLVCIFNMVPVE 556

Qy         67 -KD-TLGLPVAGQTSEFVNQVLEK 88
               || |:||||||   |  |  ||:
Db        557 RKDFTIGLPVAGIYEEVWNTELEE 580


RESULT 13
T49572
related to SHK1 KINASE-BINDING protein [imported] - Neurospora crassa
N;Alternate names: protein B2O8.200
C;Species: Neurospora crassa
C;Date: 02-Jun-2000 #sequence_revision 02-Jun-2000 #text_change 09-Jul-2004
C;Accession: T49572
R;Schulte, U.; Aign, V.; Hoheisel, J.; Brandt, P.; Fartmann, B.; Holland, R.; Nyakatura,
submitted to the Protein Sequence Database, May 2000
A;Reference number: Z25022
A;Accession: T49572
A;Status: preliminary
A;Molecule type: DNA
A;Residues: 1-718 <SCH>
A;Cross-references: UNIPROT:Q9P5Z7; UNIPARC:UPI0000069977; EMBL:AL355930; GSPDB:GN00116;
A;Experimental source: BAC clone B2O8; strain OR74A
C;Genetics:
A;Gene: NCSP:B2O8.200
A;Map position: 6
A;Introns: 45/3; 152/1; 274/1; 658/3
C;Superfamily: Schizosaccharomyces pombe negative regulator of mitosis skb1

  Query Match             13.1%;  Score 72;  DB 2;  Length 718;
  Best Local Similarity   25.7%;  Pred. No. 33;
  Matches   19;  Conservative   15;  Mismatches   30;  Indels   10;  Gaps    1;

Qy          8 RLKKKAQYEANKVKLWGLNTEYDNLLFVINSIKQEIVNRV----------QNPRDERGPN 57
              |: : |      ::|| |    :  ::::   |:|  | |          : ||    ||
Db        392 RVLRAAASTNTPIQLWALEKNQNAYVYLLRRNKREWNNSVTVIKTDMREWEGPRLAANPN 451

Qy         58 MGQKLEILIKDTLG 71
              :  ||:||: : ||
Db        452 IITKLDILVTELLG 465


RESULT 14
JN0760
H+-transporting two-sector ATPase (EC 3.6.3.14) beta chain precursor - fruit fly (Drosopl
N;Alternate names: F1-F0 ATPase beta chain
C;Species: Drosophila melanogaster
C;Date: 30-Sep-1993 #sequence_revision 20-Aug-1994 #text_change 31-Dec-2004
C;Accession: JN0760; S32130; S64699
R;Pena, P.; Garesse, R.
Biochem. Biophys. Res. Commun. 195, 785-791, 1993
A;Title: The beta subunit of the Drosophila melanogaster ATP synthase: cDNA cloning, amir
A;Reference number: JN0760; MUID:93384603; PMID:8373413
A;Accession: JN0760
A;Molecule type: mRNA
A;Residues: 1-505 <PEN>
A;Cross-references: UNIPROT:Q05825; UNIPARC:UPI000012643F
R;Garesse, R.
submitted to the EMBL Data Library, February 1993
A;Reference number: S32130
A;Accession: S32130
A;Molecule type: mRNA
A;Residues: 6-505 <GAR>
A;Cross-references: UNIPARC:UPI000016BAEB; EMBL:X71013; NID:g287944; PIDN:CAA50332.1; PII
```

```
R;Pena, P.; Ugalde, C.; Calleja, M.; Garesse, R.
Biochem. J. 312, 887-897, 1995
A;Title: Analysis of the mitochondrial ATP synthase beta-subunit gene in Drosophilidae:
er.
A;Reference number: S64699; MUID:96128076; PMID:8554535
A;Accession: S64699
A;Status: preliminary
A;Molecule type: DNA
A;Residues: 1-505 <PE2>
A;Cross-references: UNIPARC:UPI000012643F; EMBL:X86015
C;Comment: This enzyme catalyzes the synthesis of ATP coupled to H+ gradient generated i
C;Genetics:
A;Gene: FlyBase:ATPsyn-beta
A;Cross-references: FlyBase:FBgn0010217
A;Introns: 22/1; 125/3
C;Superfamily: H(+)-transporting ATP synthase; H+-transporting ATP synthase alpha chain
C;Keywords: ATP biosynthesis; hydrolase; membrane-associated complex; mitochondrion; nuc
F;1-31/Domain: transit peptide (mitochondrion) #status predicted <TNP>
F;32-505/Product: H+-transporting ATP synthase beta chain #status predicted <MAT>
F;183-190/Region: nucleotide-binding motif A (P-loop)
F;209-384/Domain: H+-transporting ATP synthase alpha chain homology <ATP>

  Query Match             12.9%;  Score 71;  DB 2;  Length 505;
  Best Local Similarity   24.8%;  Pred. No. 27;
  Matches   31;  Conservative   11;  Mismatches   21;  Indels   62;  Gaps    6;

Qy          43 IVNRVQNPRDERGP----------------NMGQKLEILIKDTLGLPV---------- 74
               |:| :  | |||||                |  : |||:    |: |
Db         121 IINVIGEPIDERGPIDTDKTAAIHAEAPEFVQMSVEQEILV---TGIKVVDLLAPYAKGG 177

Qy          75 ---------AGQT---SEFVNQVL------------EKTAEGNP------TGGLVGLRI 103
                        |:|    | :| |              |:| |||         ||:: |:
Db         178 KIGLFGGAGVGKTVLIMELINNVAKAHGGYSVFAGVGERTREGNDLYNEMIEGGVISLKD 237

Qy         104 PTSKV 108
               ||||
Db         238 KTSKV 242


RESULT 15
T22552
hypothetical protein ZK1151.1 - Caenorhabditis elegans
C;Species: Caenorhabditis elegans
C;Date: 15-Oct-1999 #sequence_revision 15-Oct-1999 #text_change 29-Oct-1999
C;Accession: T22552; T27703
R;Harris, B.
submitted to the EMBL Data Library, March 1997
A;Reference number: Z19580
A;Accession: T22552
A;Status: preliminary; translated from GB/EMBL/DDBJ
A;Molecule type: DNA
A;Residues: 1-1355 <WIL>
A;Cross-references: UNIPARC:UPI000017BCEF; EMBL:Z92788; PIDN:CAB07214.1; GSPDB:GN00019;
A;Experimental source: clone F53B8
R;Harris, B.
submitted to the EMBL Data Library, March 1997
A;Reference number: Z20408
A;Accession: T27703
A;Status: preliminary; translated from GB/EMBL/DDBJ
A;Molecule type: DNA
A;Residues: 1-1355 <WI2>
A;Cross-references: UNIPARC:UPI000017BCEF; EMBL:Z93398; PIDN:CAB07724.1; GSPDB:GN00019;
A;Experimental source: clone ZK1151
C;Genetics:
A;Gene: CESP:ZK1151.1
A;Map position: 1
A;Introns: 94/3; 124/3; 150/3; 209/3; 248/3; 311/3; 424/3; 734/2; 934/2; 1104/3; 1167/3;

  Query Match             12.9%;  Score 71;  DB 2;  Length 1355;
  Best Local Similarity   24.4%;  Pred. No. 90;
  Matches   22;  Conservative   21;  Mismatches   37;  Indels   10;  Gaps    2;

Qy          27 TEYDNLLFVINSIKQEIVNRVQNPRDERGPNMGQKLEILIKDTLGLPVAGQTSEFVNQ-- 84
                  :||:  :  :::: :| :||:: | : |     |:|::|| :   |    |  | ||:
Db        1240 SEYETEMSWLDTV-EETINRLRKPEELRPEQYQQQLDMLIAEYTNLQEHTQAIEHVNKEG 1298

Qy          85 -------VLEKTAEGNPTGGLVGLRIPTSK 107
                       :     |  : |:||:  |  |
Db        1299 GRFIHEAKIFDAKLGQYSDGIVGIHGPGIK 1328


Search completed: February 28, 2006, 08:54:21
Job time : 8.58647 secs
```

This Page Blank (uspto)

GenCore version 5.1.7

OM protein - protein search, using sw model

Run on: February 28, 2006, 08:45:31 ; Search time 39.609 Seconds
(without alignments)
1923.730 Million cell updates/sec

Title: US-10-717-665A-44_COPY_532_639
Perfect score: 551
Sequence: 1 KLAFRACRLKKKAQYEANKV.....TAEGNPTGGLVGLRIPTSKV 108

Scoring table: BLOSUM62
Gapop 10.0 , Gapext 0.5

Searched: 2166443 seqs, 705528306 residues

Total number of hits satisfying chosen parameters: 2166443

Minimum DB seq length: 0
Maximum DB seq length: 2000000000

Post-processing: Minimum Match 0%
Maximum Match 100%
Listing first 45 summaries

Database : UniProt_05.80:*
1: uniprot_sprot:*
2: uniprot_trembl:*

Pred. No. is the number of results predicted by chance to have a score greater than or equal to the score of the result being printed, and is derived by analysis of the total score distribution.

SUMMARIES

Result No.	Score	% Query Match	Length	DB	ID	Description
1	543	98.5	639	2	Q8IUR6_HUMAN	Q8iur6 homo sapien
2	536	97.3	639	2	Q8IZG1_HUMAN	Q8izg1 homo sapien
3	511	92.7	640	2	Q8CDG5_MOUSE	Q8cdg5 mus musculu
4	353	64.1	604	2	Q5HYK0_HUMAN	Q5hyk0 homo sapien
5	352.5	64.0	600	2	Q4RRX3_TETNG	Q4rrx3 tetraodon n
6	120	21.8	293	2	Q7QG21_ANOGA	Q7qg21 anopheles g
7	113.5	20.6	584	2	Q5D9Y2_SCHJA	Q5d9y2 schistosoma
8	113	20.5	755	2	Q9VC61_DROME	Q9vc61 drosophila
9	87	15.8	487	2	Q80VA1_MOUSE	Q80va1 mus musculu
10	87	15.8	818	2	Q8CID5_MOUSE	Q8cid5 mus musculu
11	87	15.8	993	2	Q6A052_MOUSE	Q6a052 mus musculu
12	87	15.8	1016	2	Q7TPS5_MOUSE	Q7tps5 mus musculu
13	85	15.4	1016	2	Q8C0U3_MOUSE	Q8c0u3 m mus muscu
14	84	15.2	1000	2	Q5RDC8_PONPY	Q5rdc8 pongo pygma
15	82.5	15.0	718	2	Q7NMF2_GLOVI	Q7nmf2 gloeobacter
16	82	14.9	571	2	Q8YV87_ANASP	Q8yv87 anabaena sp
17	82	14.9	826	2	Q7Z619_HUMAN	Q7z619 homo sapien
18	82	14.9	1000	2	Q86YS7_HUMAN	Q86ys7 homo sapien
19	82	14.9	1003	2	O60280_HUMAN	O60280 homo sapien
20	82	14.9	1005	2	Q86SU3_HUMAN	Q86su3 homo sapien
21	80	14.5	619	2	Q7ZXW0_XENLA	Q7zxw0 xenopus lae
22	79.5	14.4	534	2	Q9CAB0_ARATH	Q9cab0 arabidopsis
23	78.5	14.2	281	1	KSGA_BORBU	O51536 borrelia bu
24	78.5	14.2	428	2	Q738A9_BACC1	Q738a9 bacillus ce
25	78.5	14.2	1105	2	Q64XT2_BACFR	Q64xt2 bacteroides
26	78	14.2	876	2	Q89666_IBDV	Q89666 avian infec
27	77	14.0	3692	2	Q8KRR3_FUSNU	Q8krr3 fusobacteri
28	76.5	13.9	455	1	YP46_YEAST	Q12080 saccharomyc
29	76	13.8	467	2	Q4MZ45_THEPA	Q4mz45 theileria p
30	75.5	13.7	202	2	Q7ZVX7_BRARE	Q7zvx7 brachydanio
31	75.5	13.7	281	1	KSGA_BORGA	Q660t2 borrelia ga
32	75.5	13.7	441	2	Q4H3N6_CIOIN	Q4h3n6 ciona intes
33	75.5	13.7	577	2	Q8H0E8_LITER	Q8h0e8 lithospermu
34	75.5	13.7	642	1	GLGB_STRPN	Q97qs8 streptococc
35	75	13.6	362	2	Q4H3N5_CIOIN	Q4h3n5 ciona intes
36	75	13.6	796	2	Q73BM2_BACC1	Q73bm2 bacillus ce
37	74	13.4	386	2	Q7VS16_BORPE	Q7vs16 bordetella
38	74	13.4	386	2	Q7WEB8_BORBR	Q7web8 bordetella
39	74	13.4	701	2	Q600U1_MYCHY	Q600u1 mycoplasma
40	74	13.4	746	2	Q6U5X1_KLEPN	Q6u5x1 klebsiella
41	74	13.4	869	1	Y593_MYCLE	Q49689 mycobacteri
42	73.5	13.3	455	2	Q648B7_9ARCH	Q648b7 uncultured
43	73.5	13.3	629	2	Q4L6B5_STAHJ	Q416b5 staphylococ
44	73.5	13.3	840	2	Q4N4Q7_THEPA	Q4n4q7 theileria p
45	73.5	13.3	1104	1	CARB_ANASP	Q8yql2 anabaena sp

ALIGNMENTS

RESULT 1
Q8IUR6_HUMAN
ID Q8IUR6_HUMAN PRELIMINARY; PRT; 639 AA.
AC Q8IUR6;
DT 01-MAR-2003 (TrEMBLrel. 23, Created)
DT 01-MAR-2003 (TrEMBLrel. 23, Last sequence update)
DT 01-MAR-2004 (TrEMBLrel. 26, Last annotation update)
DE Adult retina protein.
GN Name=LOC153222;
OS Homo sapiens (Human).
OC Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;
OC Mammalia; Eutheria; Euarchontoglires; Primates; Catarrhini; Hominidae;
OC Homo.
OX NCBI_TaxID=9606;
RN [1]
RP NUCLEOTIDE SEQUENCE.
RC TISSUE=Lung;
RX MEDLINE=22388257; PubMed=12477932; DOI=10.1073/pnas.242603899;
RA Strausberg R.L., Feingold E.A., Grouse L.H., Derge J.G.,
RA Klausner R.D., Collins F.S., Wagner L., Shenmen C.M., Schuler G.D.,
RA Altschul S.F., Zeeberg B., Buetow K.H., Schaefer C.F., Bhat N.K.,
RA Hopkins R.F., Jordan H., Moore T., Max S.I., Wang J., Hsieh F.,
RA Diatchenko L., Marusina K., Farmer A.A., Rubin G.M., Hong L.,
RA Stapleton M., Soares M.B., Bonaldo M.F., Casavant T.L., Scheetz T.E.,
RA Brownstein M.J., Usdin T.B., Toshiyuki S., Carninci P., Prange C.,
RA Raha S.S., Loquellano N.A., Peters G.J., Abramson R.D., Mullahy S.J.,
RA Bosak S.A., McEwan P.J., McKernan K.J., Malek J.A., Gunaratne P.H.,
RA Richards S., Worley K.C., Hale S., Garcia A.M., Gay L.J., Hulyk S.W.,
RA Villalon D.K., Muzny D.M., Sodergren E.J., Lu X., Gibbs R.A.,
RA Fahey J., Helton E., Ketteman M., Madan A., Rodrigues S., Sanchez A.,
RA Whiting M., Madan A., Young A.C., Shevchenko Y., Bouffard G.G.,
RA Blakesley R.W., Touchman J.W., Green E.D., Dickson M.C.,
RA Rodriguez A.C., Grimwood J., Schmutz J., Myers R.M.,
RA Butterfield Y.S.N., Krzywinski M.I., Skalska U., Smailus D.E.,
RA Schnerch A., Schein J.E., Jones S.J.M., Marra M.A.;
RT "Generation and initial analysis of more than 15,000 full-length human
RT and mouse cDNA sequences.";
RL Proc. Natl. Acad. Sci. U.S.A. 99:16899-16903(2002).
RN [2]
RP NUCLEOTIDE SEQUENCE.
RC TISSUE=Lung;
RA Director MGC Project;
RL Submitted (DEC-2002) to the EMBL/GenBank/DDBJ databases.
DR EMBL; BC041709; AAH41709.1; -; mRNA.
DR Ensembl; ENSG00000164463; Homo sapiens.
DR InterPro; IPR004827; TF_bZIP.
DR PROSITE; PS00036; BZIP_BASIC; UNKNOWN_1.
SQ SEQUENCE 639 AA; 72118 MW; ECFB92D9290DEDEB CRC64;

Query Match 98.5%; Score 543; DB 2; Length 639;
Best Local Similarity 99.1%; Pred. No. 7.8e-44;
Matches 107; Conservative 0; Mismatches 1; Indels 0; Gaps 0;

Qy 1 KLAFRACRLKKKAQYEANKVKLWGLNTEYDNLLFVINSIKQEIVNRVQNPRDERGPNMGQ 60

```
           ||| ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db        532 KLASRACRLKKKAQYEANKVKLWGLNTEYDNLLFVINSIKQEIVNRVQNPRDERGPNMGQ 591

Qy         61 KLEILIKDTLGLPVAGQTSEFVNQVLEKTAEGNPTGGLVGLRIPTSKV 108
              |||||||||||||||||||||||||||||||||||||||||||||||||
Db        592 KLEILIKDTLGLPVAGQTSEFVNQVLEKTAEGNPTGGLVGLRIPTSKV 639


RESULT 2
Q8IZG1_HUMAN
ID   Q8IZG1_HUMAN PRELIMINARY;      PRT;   639 AA.
AC   Q8IZG1;
DT   01-MAR-2003 (TrEMBLrel. 23, Created)
DT   01-MAR-2003 (TrEMBLrel. 23, Last sequence update)
DT   01-MAR-2004 (TrEMBLrel. 26, Last annotation update)
DE   Adult retina protein.
OS   Homo sapiens (Human).
OC   Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;
OC   Mammalia; Eutheria; Euarchontoglires; Primates; Catarrhini; Hominidae;
OC   Homo.
OX   NCBI_TaxID=9606;
RN   [1]
RP   NUCLEOTIDE SEQUENCE.
RA   Li F., Yao K.T.;
RL   Submitted (AUG-2002) to the EMBL/GenBank/DDBJ databases.
DR   EMBL; AY139008; AAN28956.1; -; mRNA.
DR   InterPro; IPR004827; TF_bZIP.
DR   PROSITE; PS00036; BZIP_BASIC; UNKNOWN_1.
SQ   SEQUENCE   639 AA;  72090 MW;  3A19E0926B9A6406 CRC64;

  Query Match             97.3%;  Score 536;  DB 2;  Length 639;
  Best Local Similarity   98.1%;  Pred. No. 3.7e-43;
  Matches  106;  Conservative    0;  Mismatches    2;  Indels    0;  Gaps    0;

Qy          1 KLAFRACRLKKKAQYEANKVKLWGLNTEYDNLLFVINSIKQEIVNRVQNPRDERGPNMGQ 60
              ||| |||||||||||||||||||||| ||||||||||||||||||||||||||||||||||
Db        532 KLASRACRLKKKAQYEANKVKLWGLNTEYGNLLFVINSIKQEIVNRVQNPRDERGPNMGQ 591

Qy         61 KLEILIKDTLGLPVAGQTSEFVNQVLEKTAEGNPTGGLVGLRIPTSKV 108
              |||||||||||||||||||||||||||||||||||||||||||||||||
Db        592 KLEILIKDTLGLPVAGQTSEFVNQVLEKTAEGNPTGGLVGLRIPTSKV 639


RESULT 3
Q8CDG5_MOUSE
ID   Q8CDG5_MOUSE PRELIMINARY;      PRT;   640 AA.
AC   Q8CDG5;
DT   01-MAR-2003 (TrEMBLrel. 23, Created)
DT   01-MAR-2003 (TrEMBLrel. 23, Last sequence update)
DT   01-MAR-2004 (TrEMBLrel. 26, Last annotation update)
DE   Mus musculus adult male testis cDNA, RIKEN full-length enriched
DE   library, clone:4932441F15 product:hypothetical protein, full insert
DE   sequence.
GN   Name=A930001N09Rik;
OS   Mus musculus (Mouse).
OC   Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;
OC   Mammalia; Eutheria; Euarchontoglires; Glires; Rodentia; Sciurognathi;
OC   Muroidea; Muridae; Murinae; Mus.
OX   NCBI_TaxID=10090;
RN   [1]
RP   NUCLEOTIDE SEQUENCE.
RC   STRAIN=C57BL/6J; TISSUE=Testis;
RX   MEDLINE=99279253; PubMed=10349636; DOI=10.1016/S0076-6879(99)03004-9;
RA   Carninci P., Hayashizaki Y.;
RT   "High-efficiency full-length cDNA cloning.";
RL   Meth. Enzymol. 303:19-44(1999).
RN   [2]
RP   NUCLEOTIDE SEQUENCE.
RC   STRAIN=C57BL/6J; TISSUE=Testis;
RX   MEDLINE=21085660; PubMed=11217851; DOI=10.1038/35055500;
RA   Kawai J., Shinagawa A., Shibata K., Yoshino M., Itoh M., Ishii Y.,
RA   Arakawa T., Hara A., Fukunishi Y., Konno H., Adachi J., Fukuda S.,
RA   Aizawa K., Izawa M., Nishi K., Kiyosawa H., Kondo S., Yamanaka I.,
RA   Saito T., Okazaki Y., Gojobori T., Bono H., Kasukawa T., Saito R.,
RA   Kadota K., Matsuda H.A., Ashburner M., Batalov S., Casavant T.,
RA   Fleischmann W., Gaasterland T., Gissi C., King B., Kochiwa H.,
RA   Kuehl P., Lewis S., Matsuo Y., Nikaido I., Pesole G., Quackenbush J.,
RA   Schriml L.M., Staubli F., Suzuki R., Tomita M., Wagner L., Washio T.,
RA   Sakai K., Okido T., Furuno M., Aono H., Baldarelli R., Barsh G.,
RA   Blake J., Boffelli D., Bojunga N., Carninci P., de Bonaldo M.F.,
RA   Brownstein M.J., Bult C., Fletcher C., Fujita M., Gariboldi M.,
RA   Gustincich S., Hill D., Hofmann M., Hume D.A., Kamiya M., Lee N.H.,
RA   Lyons P., Marchionni L., Mashima J., Mazzarelli J., Mombaerts P.,
RA   Nordone P., Ring B., Ringwald M., Rodriguez I., Sakamoto N.,
RA   Sasaki H., Sato K., Schoenbach C., Seya T., Shibata Y., Storch K.-F.,
RA   Suzuki H., Toyo-oka K., Wang K.H., Weitz C., Whittaker C., Wilming L.,
RA   Wynshaw-Boris A., Yoshida K., Hasegawa Y., Kawaji H., Kohtsuki S.,
RA   Hayashizaki Y.;
RT   "Functional annotation of a full-length mouse cDNA collection.";
RL   Nature 409:685-690(2001).
RN   [3]
RP   NUCLEOTIDE SEQUENCE.
RC   STRAIN=C57BL/6J; TISSUE=Testis;
RA   The FANTOM Consortium,
RA   the RIKEN Genome Exploration Research Group Phase I & II Team;
RT   "Analysis of the mouse transcriptome based on functional annotation of
RT   60,770 full-length cDNAs.";
RL   Nature 420:563-573(2002).
RN   [4]
RP   NUCLEOTIDE SEQUENCE.
RC   STRAIN=C57BL/6J; TISSUE=Testis;
RX   MEDLINE=20499374; PubMed=11042159; DOI=10.1101/gr.145100;
RA   Carninci P., Shibata Y., Hayatsu N., Sugahara Y., Shibata K., Itoh M.,
RA   Konno H., Okazaki Y., Muramatsu M., Hayashizaki Y.;
RT   "Normalization and subtraction of cap-trapper-selected cDNAs to
RT   prepare full-length cDNA libraries for rapid discovery of new genes.";
RL   Genome Res. 10:1617-1630(2000).
RN   [5]
RP   NUCLEOTIDE SEQUENCE.
RC   STRAIN=C57BL/6J; TISSUE=Testis;
RX   MEDLINE=20530913; PubMed=11076861; DOI=10.1101/gr.152600;
RA   Shibata K., Itoh M., Aizawa K., Nagaoka S., Sasaki N., Carninci P.,
RA   Konno H., Akiyama J., Nishi K., Kitsunai T., Tashiro H., Itoh M.,
RA   Sumi N., Ishii Y., Nakamura S., Hazama M., Nishine T., Harada A.,
RA   Yamamoto R., Matsumoto H., Sakaguchi S., Ikegami T., Kashiwagi K.,
RA   Fujiwake S., Inoue K., Togawa Y., Izawa M., Ohara E., Watahiki M.,
RA   Yoneda Y., Ishikawa T., Ozawa K., Tanaka T., Matsuura S., Kawai J.,
RA   Okazaki Y., Muramatsu M., Inoue Y., Kira A., Hayashizaki Y.;
RT   "RIKEN integrated sequence analysis (RISA) system-384-format
RT   sequencing pipeline with 384 multicapillary sequencer.";
RL   Genome Res. 10:1757-1771(2000).
RN   [6]
RP   NUCLEOTIDE SEQUENCE.
RC   STRAIN=C57BL/6J; TISSUE=Testis;
RA   Adachi J., Aizawa K., Akimura T., Arakawa T., Bono H., Carninci P.,
RA   Fukuda S., Furuno M., Hanagaki T., Hara A., Hashizume W.,
RA   Hayashida K., Hayatsu N., Hiramoto K., Hiraoka T., Hirozane T.,
RA   Hori F., Imotani K., Ishii Y., Itoh M., Kagawa I., Kasukawa T.,
RA   Katoh H., Kawai J., Kojima Y., Kondo S., Konno H., Kouda M., Koya S.,
RA   Kurihara C., Matsuyama T., Miyazaki A., Murata M., Nakamura M.,
RA   Nishi K., Nomura K., Numazaki R., Ohno M., Ohsato N., Okazaki Y.,
RA   Saito R., Saitoh H., Sakai C., Sakai K., Sakazume N., Sano H.,
RA   Sasaki D., Shibata K., Shinagawa A., Shiraki T., Sogabe Y., Tagami M.,
RA   Tagawa A., Takahashi F., Takaku-Akahira S., Takeda Y., Tanaka T.,
RA   Tomaru A., Toya T., Yasunishi A., Muramatsu M., Hayashizaki Y.;
RL   Submitted (JUL-2001) to the EMBL/GenBank/DDBJ databases.
DR   EMBL; AK030092; BAC26779.1; -; mRNA.
DR   Ensembl; ENSMUSG00000048249; Mus musculus.
DR   MGI; MGI:1924378; A930001N09Rik.
DR   InterPro; IPR004827; TF_bZIP.
DR   PROSITE; PS00036; BZIP_BASIC; UNKNOWN_1.
KW   Hypothetical protein.
SQ   SEQUENCE   640 AA;  72598 MW;  FE02C532FA34E1DE CRC64;

  Query Match             92.7%;  Score 511;  DB 2;  Length 640;
  Best Local Similarity   93.5%;  Pred. No. 9.8e-41;
```

```
  Matches  101;  Conservative    3;  Mismatches    4;  Indels    0;  Gaps    0;

Qy          1 KLAFRACRLKKKAQYEANKVKLWGLNTEYDNLLFVINSIKQEIVNRVQNPRDERGPNMGQ 60
              ||| |||||||||||||||||||||||||||||||||||:|||||||||:|| |:|||
Db        533 KLASRACRLKKKAQYEANKVKLWGLNTEYDNLLFVINSIKQDIVNRVQNPREEREPSMGQ 592

Qy         61 KLEILIKDTLGLPVAGQTSEFVNQVLEKTAEGNPTGGLVGLRIPTSKV 108
              |||||||||||||||||||||||||| |||||||||||||||| |||
Db        593 KLEILIKDTLGLPVAGQTSEFVNQVLGKTAEGNPTGGLVGLRIPASKV 640


RESULT 4
Q5HYK0_HUMAN
ID   Q5HYK0_HUMAN PRELIMINARY;      PRT;   604 AA.
AC   Q5HYK0;
DT   10-MAY-2005 (TrEMBLrel. 30, Created)
DT   10-MAY-2005 (TrEMBLrel. 30, Last sequence update)
DT   10-MAY-2005 (TrEMBLrel. 30, Last annotation update)
DE   Hypothetical protein DKFZp313F2319 (Fragment).
GN   Name=DKFZp313F2319;
OS   Homo sapiens (Human).
OC   Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;
OC   Mammalia; Eutheria; Euarchontoglires; Primates; Catarrhini; Hominidae;
OC   Homo.
OX   NCBI_TaxID=9606;
RN   [1]
RP   NUCLEOTIDE SEQUENCE.
RC   TISSUE=Adipose;
RG   The German cDNA Consortium;
RA   Koehrer K., Beyer A., Mewes H.W., Weil B., Amid C., Osanger A.,
RA   Fobo G., Han M., Wiemann S.;
RL   Submitted (JAN-2005) to the EMBL/GenBank/DDBJ databases.
DR   EMBL; BX647573; CAI46104.1; -; mRNA.
DR   InterPro; IPR004827; TF_bZIP.
DR   PROSITE; PS00036; BZIP_BASIC; UNKNOWN_1.
KW   Hypothetical protein.
FT   NON_TER     604    604
SQ   SEQUENCE   604 AA;  68572 MW;  977C229B63E2E4C2 CRC64;

  Query Match             64.1%;  Score 353;  DB 2;  Length 604;
  Best Local Similarity   97.2%;  Pred. No. 1.8e-25;
  Matches   69;  Conservative    1;  Mismatches    1;  Indels    0;  Gaps    0;

Qy          1 KLAFRACRLKKKAQYEANKVKLWGLNTEYDNLLFVINSIKQEIVNRVQNPRDERGPNMGQ 60
              ||| ||||||||||||||||||||||||||||||||||||||||||||||||||:|||
Db        534 KLASRACRLKKKAQYEANKVKLWGLNTEYDNLLFVINSIKQEIVNRVQNPRDERGPDMGQ 593

Qy         61 KLEILIKDTLG 71
              ||||||||||
Db        594 KLEILIKDTLG 604


RESULT 5
Q4RRX3_TETNG
ID   Q4RRX3_TETNG PRELIMINARY;      PRT;   600 AA.
AC   Q4RRX3;
DT   13-SEP-2005 (TrEMBLrel. 31, Created)
DT   13-SEP-2005 (TrEMBLrel. 31, Last sequence update)
DT   13-SEP-2005 (TrEMBLrel. 31, Last annotation update)
DE   Chromosome 7 SCAF15001, whole genome shotgun sequence.
DE   (Fragment).
GN   ORFNames=GSTENG00029962001;
OS   Tetraodon nigroviridis (Green puffer).
OC   Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;
OC   Actinopterygii; Neopterygii; Teleostei; Euteleostei; Neoteleostei;
OC   Acanthomorpha; Acanthopterygii; Percomorpha; Tetraodontiformes;
OC   Tetradontoidea; Tetraodontidae; Tetraodon.
OX   NCBI_TaxID=99883;
RN   [1]
RP   NUCLEOTIDE SEQUENCE.
RA   Jaillon O., Aury J.M., Brunet F., Petit J.L., Stange-Thomann N.,
RA   Mauceli E., Bouneau L., Fischer C., Ozouf-Costaz C., Bernot A.,
RA   Nicaud S., Jaffe D., Fisher S., Lutfalla G., Dossat C., Segurens B.,
RA   Dasilva C., Salanoubat M., Levy M., Boudet N., Castellano S.,
RA   Anthouard V., Jubin C., Castelli V., Katinka M., Vacherie B.,
RA   Biemont C., Skalli Z., Cattolico L., Poulain J., De Berardinis V.,
RA   Cruaud C., Duprat S., Brottier P., Coutanceau J.P., Gouzy J.,
RA   Parra G., Lardier G., Chapple C., McKernan K.J., McEwan P., Bosak S.,
RA   Kellis M., Volff JN., Guigo R., Zody M.C., Mesirov J.,
RA   Lindblad-Toh K., Birren B., Nusbaum C., Kahn D., Robinson-Rechavi M.,
RA   Laudet V., Schachter V., Quetier F., Saurin W., Scarpelli C.,
RA   Wincker P., Lander E.S., Weissenbach J., Roest Crollius H.;
RT   "Genome duplication in the teleost fish Tetraodon nigroviridis reveals
RT   the early vertebrate proto-karyotype.";
RL   Nature 431:946-957(2004).
RN   [2]
RP   NUCLEOTIDE SEQUENCE.
RG   Genoscope; Whitehead Institute Centre for Genome Research;
RL   Submitted (FEB-2004) to the EMBL/GenBank/DDBJ databases.
CC   -!- CAUTION: The sequence shown here is derived from an
CC       EMBL/GenBank/DDBJ whole genome shotgun (WGS) entry which is
CC       preliminary data.
DR   EMBL; CAAE01015001; CAG08859.1; -; Genomic_DNA.
FT   NON_TER       1      1
FT   NON_TER     600    600
SQ   SEQUENCE   600 AA;  66291 MW;  B2F28B2F3B27B1C4 CRC64;

  Query Match             64.0%;  Score 352.5;  DB 2;  Length 600;
  Best Local Similarity   51.3%;  Pred. No. 2e-25;
  Matches   77;  Conservative   13;  Mismatches   17;  Indels   43;  Gaps    2;

Qy          1 KLAFRACRLKKKAQYEANKVKLWGLNTEYDNLLFVINSIKQEIVNRVQNPRDERGPNMGQ 60
              ||| |||||||||||||||||||:|||| |||||:||:||| ||::    |  ||
Db        452 KLASRACRLKKKAQYEANKVKLWGLSTEYDRLLFVINTIKEEIVARVED-SSPRPTNMTD 510

Qy         61 KLEILIKDTLG----------------------------------------LPVAGQT 78
              || ||: |||                                        ||||||
Db        511 TLEHLIQQTLGENPFRRDSQQNQTIFSQKRKVNDVNVDLKYILKHFSLSPQVAPPVAGQT 570

Qy         79 SEFVNQVLEKTAEGNPTGGLVGLRIPTSKV 108
              |:|||::|: |  |:|||||||:||||:
Db        571 SDFVNKILDNTGRGDPTGGLVGLRVPTSKI 600


RESULT 6
Q7QG21_ANOGA
ID   Q7QG21_ANOGA PRELIMINARY;      PRT;   293 AA.
AC   Q7QG21;
DT   01-MAR-2004 (TrEMBLrel. 26, Created)
DT   01-MAR-2004 (TrEMBLrel. 26, Last sequence update)
DT   01-MAR-2004 (TrEMBLrel. 26, Last annotation update)
DE   ENSANGP00000015170 (Fragment).
GN   ORFNames=ENSANGG00000012681;
OS   Anopheles gambiae str. PEST.
OC   Eukaryota; Metazoa; Arthropoda; Hexapoda; Insecta; Pterygota;
OC   Neoptera; Endopterygota; Diptera; Nematocera; Culicoidea; Culicidae;
OC   Anophelinae; Anopheles.
OX   NCBI_TaxID=180454;
RN   [1]
RP   NUCLEOTIDE SEQUENCE.
RC   STRAIN=PEST;
RG   The Anopheles gambiae Sequence Committee;
RT   "Anopheles gambiae re-annotation.";
RL   Submitted (APR-2002) to the EMBL/GenBank/DDBJ databases.
RN   [2]
RP   NUCLEOTIDE SEQUENCE.
RC   STRAIN=PEST;
RG   The Anopheles gambiae Sequence Committee;
RL   Submitted (APR-2004) to the EMBL/GenBank/DDBJ databases.
CC   -!- CAUTION: The sequence shown here is derived from an
CC       EMBL/GenBank/DDBJ whole genome shotgun (WGS) entry which is
CC       preliminary data.
DR   EMBL; AAAB01008844; EAA06118.2; -; Genomic_DNA.
DR   GO; GO:0005634; C:nucleus; IEA.
DR   GO; GO:0003677; F:DNA binding; IEA.
DR   InterPro; IPR004827; TF_bZIP.
```

```
DR   PROSITE; PS00036; BZIP_BASIC; UNKNOWN_1.
FT   NON_TER       1      1
FT   NON_TER     293    293
SQ   SEQUENCE   293 AA;  31535 MW;  0700521A3C593F97 CRC64;

  Query Match             21.8%;  Score 120;  DB 2;  Length 293;
  Best Local Similarity   79.3%;  Pred. No. 0.003;
  Matches   23;  Conservative    4;  Mismatches    2;  Indels    0;  Gaps    0;

Qy          1 KLAFRACRLKKKAQYEANKVKLWGLNTEY 29
              ||| |||||||||:||||:||:|| ||:
Db        247 KLASRACRLKKKAQHEANKIKLYGLETEH 275


RESULT 7
Q5D9Y2_SCHJA
ID   Q5D9Y2_SCHJA PRELIMINARY;      PRT;   584 AA.
AC   Q5D9Y2;
DT   10-MAY-2005 (TrEMBLrel. 30, Created)
DT   10-MAY-2005 (TrEMBLrel. 30, Last sequence update)
DT   10-MAY-2005 (TrEMBLrel. 30, Last annotation update)
DE   Hypothetical protein.
OS   Schistosoma japonicum (Blood fluke).
OC   Eukaryota; Metazoa; Platyhelminthes; Trematoda; Digenea; Strigeidida;
OC   Schistosomatoidea; Schistosomatidae; Schistosoma.
OX   NCBI_TaxID=6182;
RN   [1]
RP   NUCLEOTIDE SEQUENCE.
RA   Han Z.;
RT   "The full-length cDNA sequences of Schistosoma japonicum genes.";
RL   Submitted (NOV-2004) to the EMBL/GenBank/DDBJ databases.
DR   EMBL; AY815642; AAW27374.1; -; mRNA.
KW   Hypothetical protein.
SQ   SEQUENCE   584 AA;  65536 MW;  3E5055EF335323CC CRC64;

  Query Match             20.6%;  Score 113.5;  DB 2;  Length 584;
  Best Local Similarity   50.0%;  Pred. No. 0.028;
  Matches   28;  Conservative    6;  Mismatches   19;  Indels    3;  Gaps    1;

Qy          1 KLAFRACRLKKKAQYEANKVKLWGLNTEYDNLLFVINSIKQEIVNRVQNPRDERGP 56
              ||| : ||||||| :||||:|  ||  ||: |  ||  ||: |   :   ||   |
Db        345 KLASKICRLKKKAFHEANKIKYLGLEIEYNELASVIVRIKELITKYL---RDHLPP 397


RESULT 8
Q9VC61_DROME
ID   Q9VC61_DROME PRELIMINARY;      PRT;   755 AA.
AC   Q9VC61; Q8T9A9;
DT   01-MAY-2000 (TrEMBLrel. 13, Created)
DT   01-MAR-2003 (TrEMBLrel. 23, Last sequence update)
DT   10-MAY-2005 (TrEMBLrel. 30, Last annotation update)
DE   CG13624-PA, isoform A (Cg13624-pb, isoform b) (SD09792p).
GN   ORFNames=CG13624;
OS   Drosophila melanogaster (Fruit fly).
OC   Eukaryota; Metazoa; Arthropoda; Hexapoda; Insecta; Pterygota;
OC   Neoptera; Endopterygota; Diptera; Brachycera; Muscomorpha;
OC   Ephydroidea; Drosophilidae; Drosophila.
OX   NCBI_TaxID=7227;
RN   [1]
RP   NUCLEOTIDE SEQUENCE.
RX   MEDLINE=20196006; PubMed=10731132; DOI=10.1126/science.287.5461.2185;
RA   Adams M.D., Celniker S.E., Holt R.A., Evans C.A., Gocayne J.D.,
RA   Amanatides P.G., Scherer S.E., Li P.W., Hoskins R.A., Galle R.F.,
RA   George R.A., Lewis S.E., Richards S., Ashburner M., Henderson S.N.,
RA   Sutton G.G., Wortman J.R., Yandell M.D., Zhang Q., Chen L.X.,
RA   Brandon R.C., Rogers Y.-H.C., Blazej R.G., Champe M., Pfeiffer B.D.,
RA   Wan K.H., Doyle C., Baxter E.G., Helt G., Nelson C.R., Miklos G.L.G.,
RA   Abril J.F., Agbayani A., An H.-J., Andrews-Pfannkoch C., Baldwin D.,
RA   Ballew R.M., Basu A., Baxendale J., Bayraktaroglu L., Beasley E.M.,
RA   Beeson K.Y., Benos P.V., Berman B.P., Bhandari D., Bolshakov S.,
RA   Borkova D., Botchan M.R., Bouck J., Brokstein P., Brottier P.,
RA   Burtis K.C., Busam D.A., Butler H., Cadieu E., Center A., Chandra I.,
RA   Cherry J.M., Cawley S., Dahlke C., Davenport L.B., Davies P.,
RA   de Pablos B., Delcher A., Deng Z., Mays A.D., Dew I., Dietz S.M.,
RA   Dodson K., Doup L.E., Downes M., Dugan-Rocha S., Dunkov B.C., Dunn P.,
RA   Durbin K.J., Evangelista C.C., Ferraz C., Ferriera S., Fleischmann W.,
RA   Fosler C., Gabrielian A.E., Garg N.S., Gelbart W.M., Glasser K.,
RA   Glodek A., Gong F., Gorrell J.H., Gu Z., Guan P., Harris M.,
RA   Harris N.L., Harvey D.A., Heiman T.J., Hernandez J.R., Houck J.,
RA   Hostin D., Houston K.A., Howland T.J., Wei M.-H., Ibegwam C.,
RA   Jalali M., Kalush F., Karpen G.H., Ke Z., Kennison J.A., Ketchum K.A.,
RA   Kimmel B.E., Kodira C.D., Kraft C., Kravitz S., Kulp D., Lai Z.,
RA   Lasko P., Lei Y., Levitsky A.A., Li J.H., Li Z., Liang Y., Lin X.,
RA   Liu X., Mattei B., McIntosh T.C., McLeod M.P., McPherson D.,
RA   Merkulov G., Milshina N.V., Mobarry C., Morris J., Moshrefi A.,
RA   Mount S.M., Moy M., Murphy B., Murphy L., Muzny D.M., Nelson D.L.,
RA   Nelson D.R., Nelson K.A., Nixon K., Nusskern D.R., Pacleb J.M.,
RA   Palazzolo M., Pittman G.S., Pan S., Pollard J., Puri V., Reese M.G.,
RA   Reinert K., Remington K., Saunders R.D.C., Scheeler F., Shen H.,
RA   Shue B.C., Siden-Kiamos I., Simpson M., Skupski M.P., Smith T.,
RA   Spier E., Spradling A.C., Stapleton M., Strong R., Sun E.,
RA   Svirskas R., Tector C., Turner R., Venter E., Wang A.H., Wang X.,
RA   Wang Z.-Y., Wassarman D.A., Weinstock G.M., Weissenbach J.,
RA   Williams S.M., Woodage T., Worley K.C., Wu D., Yang S., Yao Q.A.,
RA   Ye J., Yeh R.-F., Zaveri J.S., Zhan M., Zhang G., Zhao Q., Zheng L.,
RA   Zheng X.H., Zhong F.N., Zhong W., Zhou X., Zhu S., Zhu X., Smith H.O.,
RA   Gibbs R.A., Myers E.W., Rubin G.M., Venter J.C.;
RT   "The genome sequence of Drosophila melanogaster.";
RL   Science 287:2185-2195(2000).
RN   [2]
RP   NUCLEOTIDE SEQUENCE.
RX   MEDLINE=22426065; PubMed=12537568;
RA   Celniker S.E., Wheeler D.A., Kronmiller B., Carlson J.W., Halpern A.,
RA   Patel S., Adams M., Champe M., Dugan S.P., Frise E., Hodgson A.,
RA   George R.A., Hoskins R.A., Laverty T., Muzny D.M., Nelson C.R.,
RA   Pacleb J.M., Park S., Pfeiffer B.D., Richards S., Sodergren E.J.,
RA   Svirskas R., Tabor P.E., Wan K., Stapleton M., Sutton G.G., Venter C.,
RA   Weinstock G., Scherer S.E., Myers E.W., Gibbs R.A., Rubin G.M.;
RT   "Finishing a whole-genome shotgun: release 3 of the Drosophila
RT   melanogaster euchromatic genome sequence.";
RL   Genome Biol. 3:RESEARCH0079-RESEARCH0079(2002).
RN   [3]
RP   NUCLEOTIDE SEQUENCE.
RX   MEDLINE=22426070; PubMed=12537573;
RA   Kaminker J.S., Bergman C.M., Kronmiller B., Carlson J.W., Svirskas R.,
RA   Patel S., Frise E., Wheeler D.A., Lewis S.E., Rubin G.M.,
RA   Ashburner M., Celniker S.E.;
RT   "The transposable elements of the Drosophila melanogaster euchromatin:
RT   a genomics perspective.";
RL   Genome Biol. 3:RESEARCH0084.1-RESEARCH0084.20(2002).
RN   [4]
RP   NUCLEOTIDE SEQUENCE.
RX   MEDLINE=22426069; PubMed=12537572;
RA   Misra S., Crosby M.A., Mungall C.J., Matthews B.B., Campbell K.S.,
RA   Hradecky P., Huang Y., Kaminker J.S., Millburn G.H., Prochnik S.E.,
RA   Smith C.D., Tupy J.L., Whitfield E.J., Bayraktaroglu L., Berman B.P.,
RA   Bettencourt B.R., Celniker S.E., de Grey A.D.N.J., Drysdale R.A.,
RA   Harris N.L., Richter J., Russo S., Schroeder A.J., Shu S.Q.,
RA   Stapleton M., Yamada C., Ashburner M., Gelbart W.M., Rubin G.M.,
RA   Lewis S.E.;
RT   "Annotation of the Drosophila melanogaster euchromatic genome: a
RT   systematic review.";
RL   Genome Biol. 3:RESEARCH0083.1-RESEARCH0083.22(2002).
RN   [5]
RP   NUCLEOTIDE SEQUENCE.
RG   Berkeley Drosophila Genome Project;
RA   Celniker S., Carlson J., Wan K., Pfeiffer B., Frise E., George R.,
RA   Hoskins R., Stapleton M., Pacleb J., Park S., Svirskas R., Smith E.,
RA   Yu C., Rubin G.;
RT   "Drosophila melanogaster release 4 sequence.";
RL   Submitted (MAR-2000) to the EMBL/GenBank/DDBJ databases.
RN   [6]
RP   NUCLEOTIDE SEQUENCE.
RG   FlyBase;
RL   Submitted (MAR-2005) to the EMBL/GenBank/DDBJ databases.
RN   [7]
```

```
RP   NUCLEOTIDE SEQUENCE.
RA   Stapleton M., Brokstein P., Hong L., Agbayani A., Carlson J.,
RA   Champe M., Chavez C., Dorsett V., Farfan D., Frise E., George R.,
RA   Gonzalez M., Guarin H., Li P., Liao G., Miranda A., Mungall C.J.,
RA   Nunoo J., Pacleb J., Paragas V., Park S., Phouanenavong S., Wan K.,
RA   Yu C., Lewis S.E., Rubin G.M., Celniker S.;
RL   Submitted (DEC-2001) to the EMBL/GenBank/DDBJ databases.
CC   -!- INTERACTION:
CC       Q9VL14:CG18619; NbExp=1; IntAct=EBI-150517, EBI-165049;
DR   EMBL; AE003748; AAF56314.2; -; Genomic_DNA.
DR   EMBL; AY069850; AAL39995.1; -; mRNA.
DR   IntAct; Q9VC61; -.
DR   Ensembl; CG13624; Drosophila melanogaster.
DR   FlyBase; FBgn0039209; CG13624.
DR   InterPro; IPR004827; TF_bZIP.
DR   PROSITE; PS00036; BZIP_BASIC; UNKNOWN_1.
SQ   SEQUENCE   755 AA;  82263 MW;  D851D42235FF10E5 CRC64;

  Query Match             20.5%;  Score 113;  DB 2;  Length 755;
  Best Local Similarity   75.9%;  Pred. No. 0.041;
  Matches   22;  Conservative    4;  Mismatches    3;  Indels    0;  Gaps    0;

Qy          1 KLAFRACRLKKKAQYEANKVKLWGLNTEY 29
              ||| |||||||||:||||:||:||  |:
Db        708 KLASRACRLKKKAQHEANKIKLFGLEIEH 736


RESULT 9
Q80VA1_MOUSE
ID   Q80VA1_MOUSE PRELIMINARY;      PRT;   487 AA.
AC   Q80VA1;
DT   01-JUN-2003 (TrEMBLrel. 24, Created)
DT   01-JUN-2003 (TrEMBLrel. 24, Last sequence update)
DT   01-MAR-2004 (TrEMBLrel. 26, Last annotation update)
DE   5730419I09Rik protein (Fragment).
GN   Name=5730419I09Rik;
OS   Mus musculus (Mouse).
OC   Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;
OC   Mammalia; Eutheria; Euarchontoglires; Glires; Rodentia; Sciurognathi;
OC   Muroidea; Muridae; Murinae; Mus.
OX   NCBI_TaxID=10090;
RN   [1]
RP   NUCLEOTIDE SEQUENCE.
RC   STRAIN=FVB/N;
RC   TISSUE=Mammary tumor. Metallothionien-TGF alpha model. 10 month old
RC   virgin mouse. Taken by biopsy.;
RX   MEDLINE=22388257; PubMed=12477932; DOI=10.1073/pnas.242603899;
RA   Strausberg R.L., Feingold E.A., Grouse L.H., Derge J.G.,
RA   Klausner R.D., Collins F.S., Wagner L., Shenmen C.M., Schuler G.D.,
RA   Altschul S.F., Zeeberg B., Buetow K.H., Schaefer C.F., Bhat N.K.,
RA   Hopkins R.F., Jordan H., Moore T., Max S.I., Wang J., Hsieh F.,
RA   Diatchenko L., Marusina K., Farmer A.A., Rubin G.M., Hong L.,
RA   Stapleton M., Soares M.B., Bonaldo M.F., Casavant T.L., Scheetz T.E.,
RA   Brownstein M.J., Usdin T.B., Toshiyuki S., Carninci P., Prange C.,
RA   Raha S.S., Loquellano N.A., Peters G.J., Abramson R.D., Mullahy S.J.,
RA   Bosak S.A., McEwan P.J., McKernan K.J., Malek J.A., Gunaratne P.H.,
RA   Richards S., Worley K.C., Hale S., Garcia A.M., Gay L.J., Hulyk S.W.,
RA   Villalon D.K., Muzny D.M., Sodergren E.J., Lu X., Gibbs R.A.,
RA   Fahey J., Helton E., Ketteman M., Madan A., Rodrigues S., Sanchez A.,
RA   Whiting M., Madan A., Young A.C., Shevchenko Y., Bouffard G.G.,
RA   Blakesley R.W., Touchman J.W., Green E.D., Dickson M.C.,
RA   Rodriguez A.C., Grimwood J., Schmutz J., Myers R.M.,
RA   Butterfield Y.S.N., Krzywinski M.I., Skalska U., Smailus D.E.,
RA   Schnerch A., Schein J.E., Jones S.J.M., Marra M.A.;
RT   "Generation and initial analysis of more than 15,000 full-length human
RT   and mouse cDNA sequences.";
RL   Proc. Natl. Acad. Sci. U.S.A. 99:16899-16903(2002).
RN   [2]
RP   NUCLEOTIDE SEQUENCE.
RC   STRAIN=FVB/N;
RC   TISSUE=Mammary tumor. Metallothionien-TGF alpha model. 10 month old
RC   virgin mouse. Taken by biopsy.;
RA   Strausberg R.;
RL   Submitted (MAR-2003) to the EMBL/GenBank/DDBJ databases.
DR   EMBL; BC049905; AAH49905.1; -; mRNA.
DR   Ensembl; ENSMUSG00000030279; Mus musculus.
DR   MGI; MGI:1921991; 5730419I09Rik.
FT   NON_TER       1      1
SQ   SEQUENCE   487 AA;  53277 MW;  9CE3489ADD52AAC7 CRC64;

  Query Match             15.8%;  Score 87;  DB 2;  Length 487;
  Best Local Similarity   29.4%;  Pred. No. 8.3;
  Matches   30;  Conservative   22;  Mismatches   26;  Indels   24;  Gaps    6;

Qy          5 RACRLKKKAQYEANKVKLWGLNTEYDNLL-FVINSIKQEIVNRVQNPRDERGPN--MGQK 61
              | |||||||| |||          |   ||| |:   :  :::|:::     :| |   | :
Db          5 RLCRLKKKAQAEANA-------TAISNLLPFMEYEVHTQLMNKLK----LKGMNALFGLR 53

Qy         62 LEILIKDTLGLPVAGQTSEFVNQVLEKTAEGNPTGGLVGLRI 103
              ::| : :|: : :|  |  ::  :        || |  |::|
Db         54 IQITVGETMLMGLASATGVYLAAL--------PTPG--GIQI 85


RESULT 10
Q8CID5_MOUSE
ID   Q8CID5_MOUSE PRELIMINARY;      PRT;   818 AA.
AC   Q8CID5;
DT   01-MAR-2003 (TrEMBLrel. 23, Created)
DT   01-MAR-2003 (TrEMBLrel. 23, Last sequence update)
DT   01-MAR-2004 (TrEMBLrel. 26, Last annotation update)
DE   RIKEN cDNA 5730419I09.
GN   Name=5730419I09Rik;
OS   Mus musculus (Mouse).
OC   Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;
OC   Mammalia; Eutheria; Euarchontoglires; Glires; Rodentia; Sciurognathi;
OC   Muroidea; Muridae; Murinae; Mus.
OX   NCBI_TaxID=10090;
RN   [1]
RP   NUCLEOTIDE SEQUENCE.
RC   STRAIN=FVB/N; TISSUE=Mammary tumor. C3;
RX   MEDLINE=22388257; PubMed=12477932; DOI=10.1073/pnas.242603899;
RA   Strausberg R.L., Feingold E.A., Grouse L.H., Derge J.G.,
RA   Klausner R.D., Collins F.S., Wagner L., Shenmen C.M., Schuler G.D.,
RA   Altschul S.F., Zeeberg B., Buetow K.H., Schaefer C.F., Bhat N.K.,
RA   Hopkins R.F., Jordan H., Moore T., Max S.I., Wang J., Hsieh F.,
RA   Diatchenko L., Marusina K., Farmer A.A., Rubin G.M., Hong L.,
RA   Stapleton M., Soares M.B., Bonaldo M.F., Casavant T.L., Scheetz T.E.,
RA   Brownstein M.J., Usdin T.B., Toshiyuki S., Carninci P., Prange C.,
RA   Raha S.S., Loquellano N.A., Peters G.J., Abramson R.D., Mullahy S.J.,
RA   Bosak S.A., McEwan P.J., McKernan K.J., Malek J.A., Gunaratne P.H.,
RA   Richards S., Worley K.C., Hale S., Garcia A.M., Gay L.J., Hulyk S.W.,
RA   Villalon D.K., Muzny D.M., Sodergren E.J., Lu X., Gibbs R.A.,
RA   Fahey J., Helton E., Ketteman M., Madan A., Rodrigues S., Sanchez A.,
RA   Whiting M., Madan A., Young A.C., Shevchenko Y., Bouffard G.G.,
RA   Blakesley R.W., Touchman J.W., Green E.D., Dickson M.C.,
RA   Rodriguez A.C., Grimwood J., Schmutz J., Myers R.M.,
RA   Butterfield Y.S.N., Krzywinski M.I., Skalska U., Smailus D.E.,
RA   Schnerch A., Schein J.E., Jones S.J.M., Marra M.A.;
RT   "Generation and initial analysis of more than 15,000 full-length human
RT   and mouse cDNA sequences.";
RL   Proc. Natl. Acad. Sci. U.S.A. 99:16899-16903(2002).
RN   [2]
RP   NUCLEOTIDE SEQUENCE.
RC   STRAIN=FVB/N; TISSUE=Mammary tumor. C3;
RA   Director MGC Project;
RL   Submitted (APR-2002) to the EMBL/GenBank/DDBJ databases.
DR   EMBL; BC027763; AAH27763.1; -; mRNA.
DR   Ensembl; ENSMUSG00000030279; Mus musculus.
DR   MGI; MGI:1921991; 5730419I09Rik.
SQ   SEQUENCE   818 AA;  88960 MW;  CAD26DC93A6204D8 CRC64;

  Query Match             15.8%;  Score 87;  DB 2;  Length 818;
  Best Local Similarity   29.4%;  Pred. No. 15;
  Matches   30;  Conservative   22;  Mismatches   26;  Indels   24;  Gaps    6;

Qy          5 RACRLKKKAQYEANKVKLWGLNTEYDNLL-FVINSIKQEIVNRVQNPRDERGPN--MGQK 61
```

```
           | |||||||| |||        |   ||| |:   :  :::|:::     :| |   | :
Db        319 RLCRLKKKAQAEANA-------TAISNLLPFMEYEVHTQLMNKLK----LKGMNALFGLR 367

Qy         62 LEILIKDTLGLPVAGQTSEFVNQVLEKTAEGNPTGGLVGLRI 103
              ::| : :|: : :|  |  ::  :         || |  |::|
Db        368 IQITVGETMLMGLASATGVYLAAL--------PTPG--GIQI 399


RESULT 11
Q6A052_MOUSE
ID   Q6A052_MOUSE PRELIMINARY;      PRT;   993 AA.
AC   Q6A052;
DT   25-OCT-2004 (TrEMBLrel. 28, Created)
DT   25-OCT-2004 (TrEMBLrel. 28, Last sequence update)
DT   25-OCT-2004 (TrEMBLrel. 28, Last annotation update)
DE   MKIAA0528 protein (Fragment).
GN   Name=5730419I09Rik; Synonyms=mKIAA0528;
OS   Mus musculus (Mouse).
OC   Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;
OC   Mammalia; Eutheria; Euarchontoglires; Glires; Rodentia; Sciurognathi;
OC   Muroidea; Muridae; Murinae; Mus.
OX   NCBI_TaxID=10090;
RN   [1]
RP   NUCLEOTIDE SEQUENCE.
RC   TISSUE=Fetal brain;
RX   PubMed=15368895;
RA   Okazaki N., Kikuno R., Ohara R., Inamoto S., Koseki H., Hiraoka S.,
RA   Saga Y., Seino S., Nishimura M., Kaisho T., Hoshino K., Kitamura H.,
RA   Nagase T., Ohara O., Koga H.;
RT   "Prediction of the coding sequences of mouse homologues of KIAA gene:
RT   IV. The complete nucleotide sequences of 500 mouse KIAA-homologous
RT   cDNAs identified by screening of terminal sequences of cDNA clones
RT   randomly sampled from size-fractionated libraries.";
RL   DNA Res. 11:205-218(2004).
CC   -!- SIMILARITY: Contains 1 C2 domain.
DR   EMBL; AK172966; BAD32244.1; -; mRNA.
DR   Ensembl; ENSMUSG00000030279; Mus musculus.
DR   MGI; MGI:1921991; 5730419I09Rik.
DR   GO; GO:0016020; C:membrane; IEA.
DR   GO; GO:0008021; C:synaptic vesicle; IEA.
DR   GO; GO:0005215; F:transporter activity; IEA.
DR   GO; GO:0006810; P:transport; IEA.
DR   InterPro; IPR000008; C2.
DR   InterPro; IPR001565; Synaptotagmin.
DR   Pfam; PF00168; C2; 1.
DR   PRINTS; PR00360; C2DOMAIN.
DR   PRINTS; PR00399; SYNAPTOTAGMN.
DR   SMART; SM00239; C2; 1.
DR   PROSITE; PS50004; C2_DOMAIN; 1.
FT   NON_TER       1       1
SQ   SEQUENCE   993 AA;  109096 MW;  9C7853BC7E8085C6 CRC64;

  Query Match             15.8%;  Score 87;  DB 2;  Length 993;
  Best Local Similarity   29.4%;  Pred. No. 18;
  Matches   30;  Conservative   22;  Mismatches   26;  Indels   24;  Gaps    6;

Qy          5 RACRLKKKAQYEANKVKLWGLNTEYDNLL-FVINSIKQEIVNRVQNPRDERGPN--MGQK 61
              | |||||||| |||        |   ||| |:   :  :::|:::     :| |   | :
Db        511 RLCRLKKKAQAEANA-------TAISNLLPFMEYEVHTQLMNKLK----LKGMNALFGLR 559

Qy         62 LEILIKDTLGLPVAGQTSEFVNQVLEKTAEGNPTGGLVGLRI 103
              ::| : :|: : :|  |  ::  :         || |  |::|
Db        560 IQITVGETMLMGLASATGVYLAAL--------PTPG--GIQI 591


RESULT 12
Q7TPS5_MOUSE
ID   Q7TPS5_MOUSE PRELIMINARY;      PRT;  1016 AA.
AC   Q7TPS5;
DT   01-OCT-2003 (TrEMBLrel. 25, Created)
DT   01-OCT-2003 (TrEMBLrel. 25, Last sequence update)
DT   01-MAR-2004 (TrEMBLrel. 26, Last annotation update)
DE   5730419I09Rik protein.
GN   Name=5730419I09Rik;
OS   Mus musculus (Mouse).
OC   Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;
OC   Mammalia; Eutheria; Euarchontoglires; Glires; Rodentia; Sciurognathi;
OC   Muroidea; Muridae; Murinae; Mus.
OX   NCBI_TaxID=10090;
RN   [1]
RP   NUCLEOTIDE SEQUENCE.
RC   STRAIN=C57BL/6J; TISSUE=Egg;
RX   MEDLINE=22388257; PubMed=12477932; DOI=10.1073/pnas.242603899;
RA   Strausberg R.L., Feingold E.A., Grouse L.H., Derge J.G.,
RA   Klausner R.D., Collins F.S., Wagner L., Shenmen C.M., Schuler G.D.,
RA   Altschul S.F., Zeeberg B., Buetow K.H., Schaefer C.F., Bhat N.K.,
RA   Hopkins R.F., Jordan H., Moore T., Max S.I., Wang J., Hsieh F.,
RA   Diatchenko L., Marusina K., Farmer A.A., Rubin G.M., Hong L.,
RA   Stapleton M., Soares M.B., Bonaldo M.F., Casavant T.L., Scheetz T.E.,
RA   Brownstein M.J., Usdin T.B., Toshiyuki S., Carninci P., Prange C.,
RA   Raha S.S., Loquellano N.A., Peters G.J., Abramson R.D., Mullahy S.J.,
RA   Bosak S.A., McEwan P.J., McKernan K.J., Malek J.A., Gunaratne P.H.,
RA   Richards S., Worley K.C., Hale S., Garcia A.M., Gay L.J., Hulyk S.W.,
RA   Villalon D.K., Muzny D.M., Sodergren E.J., Lu X., Gibbs R.A.,
RA   Fahey J., Helton E., Ketteman M., Madan A., Rodrigues S., Sanchez A.,
RA   Whiting M., Madan A., Young A.C., Shevchenko Y., Bouffard G.G.,
RA   Blakesley R.W., Touchman J.W., Green E.D., Dickson M.C.,
RA   Rodriguez A.C., Grimwood J., Schmutz J., Myers R.M.,
RA   Butterfield Y.S.N., Krzywinski M.I., Skalska U., Smailus D.E.,
RA   Schnerch A., Schein J.E., Jones S.J.M., Marra M.A.;
RT   "Generation and initial analysis of more than 15,000 full-length human
RT   and mouse cDNA sequences.";
RL   Proc. Natl. Acad. Sci. U.S.A. 99:16899-16903(2002).
RN   [2]
RP   NUCLEOTIDE SEQUENCE.
RC   STRAIN=C57BL/6J; TISSUE=Egg;
RA   Strausberg R.;
RL   Submitted (JUN-2003) to the EMBL/GenBank/DDBJ databases.
CC   -!- SIMILARITY: Contains 1 C2 domain.
DR   EMBL; BC053913; AAH53913.1; -; mRNA.
DR   HSSP; P21707; 1K5W.
DR   Ensembl; ENSMUSG00000030279; Mus musculus.
DR   MGI; MGI:1921991; 5730419I09Rik.
DR   GO; GO:0016020; C:membrane; IEA.
DR   GO; GO:0008021; C:synaptic vesicle; IEA.
DR   GO; GO:0005215; F:transporter activity; IEA.
DR   GO; GO:0006810; P:transport; IEA.
DR   InterPro; IPR000008; C2.
DR   InterPro; IPR001565; Synaptotagmin.
DR   Pfam; PF00168; C2; 1.
DR   PRINTS; PR00360; C2DOMAIN.
DR   PRINTS; PR00399; SYNAPTOTAGMN.
DR   SMART; SM00239; C2; 1.
DR   PROSITE; PS50004; C2_DOMAIN; 1.
SQ   SEQUENCE   1016 AA;  111692 MW;  F751F599FE6016FB CRC64;

  Query Match             15.8%;  Score 87;  DB 2;  Length 1016;
  Best Local Similarity   29.4%;  Pred. No. 19;
  Matches   30;  Conservative   22;  Mismatches   26;  Indels   24;  Gaps    6;

Qy          5 RACRLKKKAQYEANKVKLWGLNTEYDNLL-FVINSIKQEIVNRVQNPRDERGPN--MGQK 61
              | |||||||| |||        |   ||| |:   :  :::|:::     :| |   | :
Db        517 RLCRLKKKAQAEANA-------TAISNLLPFMEYEVHTQLMNKLK----LKGMNALFGLR 565

Qy         62 LEILIKDTLGLPVAGQTSEFVNQVLEKTAEGNPTGGLVGLRI 103
              ::| : :|: : :|  |  ::  :         || |  |::|
Db        566 IQITVGETMLMGLASATGVYLAAL--------PTPG--GIQI 597


RESULT 13
Q8C0U3_MOUSE
ID   Q8C0U3_MOUSE PRELIMINARY;      PRT;  1016 AA.
AC   Q8C0U3;
DT   01-MAR-2003 (TrEMBLrel. 23, Created)
DT   01-MAR-2003 (TrEMBLrel. 23, Last sequence update)
DT   01-MAR-2004 (TrEMBLrel. 26, Last annotation update)
```

```
DE   Mus musculus adult male testis cDNA, RIKEN full-length enriched
DE   library, clone:4930598L16 product:hypothetical C2 domain/C2-domain
DE   profile/Synaptotagmin/Cytochrome c family heme-binding site containing
DE   protein, full insert sequence.
GN   Name=5730419I09Rik;
OS   Mus musculus (Mouse).
OC   Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;
OC   Mammalia; Eutheria; Euarchontoglires; Glires; Rodentia; Sciurognathi;
OC   Muroidea; Muridae; Murinae; Mus.
OX   NCBI_TaxID=10090;
RN   [1]
RP   NUCLEOTIDE SEQUENCE.
RC   STRAIN=C57BL/6J; TISSUE=Testis;
RX   MEDLINE=99279253; PubMed=10349636; DOI=10.1016/S0076-6879(99)03004-9;
RA   Carninci P., Hayashizaki Y.;
RT   "High-efficiency full-length cDNA cloning.";
RL   Meth. Enzymol. 303:19-44(1999).
RN   [2]
RP   NUCLEOTIDE SEQUENCE.
RC   STRAIN=C57BL/6J; TISSUE=Testis;
RX   MEDLINE=21085660; PubMed=11217851; DOI=10.1038/35055500;
RA   Kawai J., Shinagawa A., Shibata K., Yoshino M., Itoh M., Ishii Y.,
RA   Arakawa T., Hara A., Fukunishi Y., Konno H., Adachi J., Fukuda S.,
RA   Aizawa K., Izawa M., Nishi K., Kiyosawa H., Kondo S., Yamanaka I.,
RA   Saito T., Okazaki Y., Gojobori T., Bono H., Kasukawa T., Saito R.,
RA   Kadota K., Matsuda H.A., Ashburner M., Batalov S., Casavant T.,
RA   Fleischmann W., Gaasterland T., Gissi C., King B., Kochiwa H.,
RA   Kuehl P., Lewis S., Matsuo Y., Nikaido I., Pesole G., Quackenbush J.,
RA   Schriml L.M., Staubli F., Suzuki R., Tomita M., Wagner L., Washio T.,
RA   Sakai K., Okido T., Furuno M., Aono H., Baldarelli R., Barsh G.,
RA   Blake J., Boffelli D., Bojunga N., Carninci P., de Bonaldo M.F.,
RA   Brownstein M.J., Bult C., Fletcher C., Fujita M., Gariboldi M.,
RA   Gustincich S., Hill D., Hofmann M., Hume D.A., Kamiya M., Lee N.H.,
RA   Lyons P., Marchionni L., Mashima J., Mazzarelli J., Mombaerts P.,
RA   Nordone P., Ring B., Ringwald M., Rodriguez I., Sakamoto N.,
RA   Sasaki H., Sato K., Schoenbach C., Seya T., Shibata Y., Storch K.-F.,
RA   Suzuki H., Toyo-oka K., Wang K.H., Weitz C., Whittaker C., Wilming L.,
RA   Wynshaw-Boris A., Yoshida K., Hasegawa Y., Kawaji H., Kohtsuki S.,
RA   Hayashizaki Y.;
RT   "Functional annotation of a full-length mouse cDNA collection.";
RL   Nature 409:685-690(2001).
RN   [3]
RP   NUCLEOTIDE SEQUENCE.
RC   STRAIN=C57BL/6J; TISSUE=Testis;
RA   The FANTOM Consortium,
RA   the RIKEN Genome Exploration Research Group Phase I & II Team;
RT   "Analysis of the mouse transcriptome based on functional annotation of
RT   60,770 full-length cDNAs.";
RL   Nature 420:563-573(2002).
RN   [4]
RP   NUCLEOTIDE SEQUENCE.
RC   STRAIN=C57BL/6J; TISSUE=Testis;
RX   MEDLINE=20499374; PubMed=11042159; DOI=10.1101/gr.145100;
RA   Carninci P., Shibata Y., Hayatsu N., Sugahara Y., Shibata K., Itoh M.,
RA   Konno H., Okazaki Y., Muramatsu M., Hayashizaki Y.;
RT   "Normalization and subtraction of cap-trapper-selected cDNAs to
RT   prepare full-length cDNA libraries for rapid discovery of new genes.";
RL   Genome Res. 10:1617-1630(2000).
RN   [5]
RP   NUCLEOTIDE SEQUENCE.
RC   STRAIN=C57BL/6J; TISSUE=Testis;
RX   MEDLINE=20530913; PubMed=11076861; DOI=10.1101/gr.152600;
RA   Shibata K., Itoh M., Aizawa K., Nagaoka S., Sasaki N., Carninci P.,
RA   Konno H., Akiyama J., Nishi K., Kitsunai T., Tashiro H., Itoh M.,
RA   Sumi N., Ishii Y., Nakamura S., Hazama M., Nishine T., Harada A.,
RA   Yamamoto R., Matsumoto H., Sakaguchi S., Ikegami T., Kashiwagi K.,
RA   Fujiwake S., Inoue K., Togawa Y., Izawa M., Ohara E., Watahiki M.,
RA   Yoneda Y., Ishikawa T., Ozawa K., Tanaka T., Matsuura S., Kawai J.,
RA   Okazaki Y., Muramatsu M., Inoue Y., Kira A., Hayashizaki Y.;
RT   "RIKEN integrated sequence analysis (RISA) system-384-format
RT   sequencing pipeline with 384 multicapillary sequencer.";
RL   Genome Res. 10:1757-1771(2000).
RN   [6]
RP   NUCLEOTIDE SEQUENCE.
RC   STRAIN=C57BL/6J; TISSUE=Testis;
RA   Adachi J., Aizawa K., Akimura T., Arakawa T., Bono H., Carninci P.,
RA   Fukuda S., Furuno M., Hanagaki T., Hara A., Hashizume W.,
RA   Hayashida K., Hayatsu N., Hiramoto K., Hiraoka T., Hirozane T.,
RA   Hori F., Imotani K., Ishii Y., Itoh M., Kagawa I., Kasukawa T.,
RA   Katoh H., Kawai J., Kojima Y., Kondo S., Konno H., Kouda M., Koya S.,
RA   Kurihara C., Matsuyama T., Miyazaki A., Murata M., Nakamura M.,
RA   Nishi K., Nomura K., Numazaki R., Ohno M., Ohsato N., Okazaki Y.,
RA   Saito R., Saitoh H., Sakai C., Sakai K., Sakazume N., Sano H.,
RA   Sasaki D., Shibata K., Shinagawa A., Shiraki T., Sogabe Y., Tagami M.,
RA   Tagawa A., Takahashi F., Takaku-Akahira S., Takeda Y., Tanaka T.,
RA   Tomaru A., Toya T., Yasunishi A., Muramatsu M., Hayashizaki Y.;
RL   Submitted (JUL-2001) to the EMBL/GenBank/DDBJ databases.
CC   -!- SIMILARITY: Contains 1 C2 domain.
DR   EMBL; AK029825; BAC26634.1; -; mRNA.
DR   HSSP; P47709; 3RPB.
DR   Ensembl; ENSMUSG00000030279; Mus musculus.
DR   MGI; MGI:1921991; 5730419I09Rik.
DR   GO; GO:0016020; C:membrane; IEA.
DR   GO; GO:0008021; C:synaptic vesicle; IEA.
DR   GO; GO:0005215; F:transporter activity; IEA.
DR   GO; GO:0006810; P:transport; IEA.
DR   InterPro; IPR000008; C2.
DR   InterPro; IPR001565; Synaptotagmin.
DR   Pfam; PF00168; C2; 1.
DR   PRINTS; PR00360; C2DOMAIN.
DR   PRINTS; PR00399; SYNAPTOTAGMN.
DR   SMART; SM00239; C2; 1.
DR   PROSITE; PS50004; C2_DOMAIN; 1.
KW   Hypothetical protein.
SQ   SEQUENCE   1016 AA;  111684 MW;  CA8958E385587E0B CRC64;

  Query Match             15.4%;  Score 85;  DB 2;  Length 1016;
  Best Local Similarity   30.9%;  Pred. No. 30;
  Matches   25;  Conservative   19;  Mismatches   23;  Indels   14;  Gaps    4;

Qy          5 RACRLKKKAQYEANKVKLWGLNTEYDNLL-FVINSIKQEIVNRVQNPRDERGPN--MGQK 61
              | |||||||| |||       |   ||| |:   :  :::|:::     :| |   | :
Db        517 RLCRLKKKAQAEANA-------TAISNLLPFMEYEVHTQLMNKLK----LKGMNALFGLR 565

Qy         62 LEILIKDTLGLPVAGQTSEFV 82
              ::| : :|: : :|  |  ::
Db        566 IQITVGETMLMGLASATGVYL 586


RESULT 14
Q5RDC8_PONPY
ID   Q5RDC8_PONPY PRELIMINARY;      PRT;  1000 AA.
AC   Q5RDC8;
DT   01-FEB-2005 (TrEMBLrel. 29, Created)
DT   01-FEB-2005 (TrEMBLrel. 29, Last sequence update)
DT   01-FEB-2005 (TrEMBLrel. 29, Last annotation update)
DE   Hypothetical protein DKFZp469O0513.
GN   Name=DKFZp469O0513;
OS   Pongo pygmaeus (Orangutan).
OC   Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;
OC   Mammalia; Eutheria; Euarchontoglires; Primates; Catarrhini; Hominidae;
OC   Pongo.
OX   NCBI_TaxID=9600;
RN   [1]
RP   NUCLEOTIDE SEQUENCE.
RC   TISSUE=Kidney;
RG   The German cDNA Consortium;
RA   Poustka A., Albert R., Moosmayer P., Schupp I., Wellenreuther R.,
RA   Mewes H.W., Weil B., Amid C., Osanger A., Fobo G., Han M., Wiemann S.;
RL   Submitted (NOV-2004) to the EMBL/GenBank/DDBJ databases.
CC   -!- SIMILARITY: Contains 1 C2 domain.
DR   EMBL; CR857986; CAH90229.1; -; mRNA.
DR   GO; GO:0016020; C:membrane; IEA.
DR   GO; GO:0008021; C:synaptic vesicle; IEA.
DR   GO; GO:0005215; F:transporter activity; IEA.
DR   GO; GO:0006810; P:transport; IEA.
```

```
DR   InterPro; IPR000008; C2.
DR   InterPro; IPR001565; Synaptotagmin.
DR   Pfam; PF00168; C2; 1.
DR   PRINTS; PR00360; C2DOMAIN.
DR   PRINTS; PR00399; SYNAPTOTAGMN.
DR   SMART; SM00239; C2; 1.
DR   PROSITE; PS50004; C2_DOMAIN_2; 1.
KW   Hypothetical protein.
SQ   SEQUENCE   1000 AA;  110365 MW;  FB307C795DA63C92 CRC64;

  Query Match             15.2%;  Score 84;  DB 2;  Length 1000;
  Best Local Similarity   29.7%;  Pred. No. 36;
  Matches   30;  Conservative   18;  Mismatches   31;  Indels   22;  Gaps    5;

Qy          5 RACRLKKKAQYEANKVKLWGLNTEYDNLLFVINSIKQEIVNRVQNPRDERGPN--MGQKL 62
              | |||||||| |||        |   |||     |: |:  :: |    :| |   | ::
Db        517 RLCRLKKKAQAEANA-------TAISNLL---PFIEYEVHTQLMNKLKLKGMNALFGLRI 566

Qy         63 EILIKDTLGLPVAGQTSEFVNQVLEKTAEGNPTGGLVGLRI 103
              :| : : : : :|   |  ::  :        || |  |::|
Db        567 QITVGENMLMGLASATGVYLAAL--------PTPG--GIQI 597


RESULT 15
Q7NMF2_GLOVI
ID   Q7NMF2_GLOVI PRELIMINARY;      PRT;   718 AA.
AC   Q7NMF2;
DT   01-MAR-2004 (TrEMBLrel. 26, Created)
DT   01-MAR-2004 (TrEMBLrel. 26, Last sequence update)
DT   01-MAR-2004 (TrEMBLrel. 26, Last annotation update)
DE   Two-component sensor histidine kinase.
GN   OrderedLocusNames=gll0814;
OS   Gloeobacter violaceus.
OC   Bacteria; Cyanobacteria; Gloeobacteria; Gloeobacterales; Gloeobacter.
OX   NCBI_TaxID=33072;
RN   [1]
RP   NUCLEOTIDE SEQUENCE [LARGE SCALE GENOMIC DNA].
RC   STRAIN=PCC 7421;
RX   MEDLINE=22977040; PubMed=14621292;
RA   Nakamura Y., Kaneko T., Sato S., Mimuro M., Miyashita H., Tsuchiya T.,
RA   Sasamoto S., Watanabe A., Kawashima K., Kishida Y., Kiyokawa C.,
RA   Kohara M., Matsumoto M., Matsuno A., Nakazaki N., Shimpo S.,
RA   Takeuchi C., Yamada M., Tabata S.;
RT   "Complete genome structure of Gloeobacter violaceus PCC 7421, a
RT   cyanobacterium that lacks thylakoids.";
RL   DNA Res. 10:137-145(2003).
DR   EMBL; BA000045; BAC88755.1; -; Genomic_DNA.
DR   GO; GO:0016020; C:membrane; IEA.
DR   GO; GO:0005524; F:ATP binding; IEA.
DR   GO; GO:0016301; F:kinase activity; IEA.
DR   GO; GO:0000155; F:two-component sensor molecule activity; IEA.
DR   GO; GO:0006468; P:protein amino acid phosphorylation; IEA.
DR   GO; GO:0000160; P:two-component signal transduction system (p. . .; IEA.
DR   InterPro; IPR003594; ATPbind_ATPase.
DR   InterPro; IPR005467; His_kinase.
DR   InterPro; IPR003661; His_kinA_N.
DR   InterPro; IPR003660; His_kin_HAMP.
DR   InterPro; IPR004358; His_kin_like_C.
DR   InterPro; IPR000014; PAS.
DR   InterPro; IPR000700; PAS-assoc_C.
DR   Pfam; PF00672; HAMP; 1.
DR   Pfam; PF02518; HATPase_c; 1.
DR   Pfam; PF00512; HisKA; 1.
DR   Pfam; PF00989; PAS; 1.
DR   PRINTS; PR00344; BCTRLSENSOR.
DR   TIGRFAMs; TIGR00229; sensory_box; 1.
DR   PROSITE; PS50109; HIS_KIN; 1.
DR   PROSITE; PS50113; PAC; 1.
KW   Complete proteome; Kinase.
SQ   SEQUENCE   718 AA;  78094 MW;  D3E695454C0956DC CRC64;

  Query Match             15.0%;  Score 82.5;  DB 2;  Length 718;
  Best Local Similarity   24.3%;  Pred. No. 35;
  Matches   36;  Conservative   20;  Mismatches   45;  Indels   47;  Gaps    4;

Qy          3 AFRACRLKKKAQYEANKVKLWGLNTEYDNLLFVINSIKQEIV----------NR----- 46
              :| | | :  |: |||:  |  |  | | |  || :| : ::          ||
Db        345 SFGAMRERIGAEAEANRRLLANLTAEKDKLELVIEAIAEGVLVYDSSGRLRTANRALWSL 404

Qy         47 VQNPRDERGPNMGQKLEILIKDTLGLPVAGQTSEFVNQVLE---------------- 87
              : :| |    :     : ::| || ||| : : |   | |
Db        405 LDSPPGE----LAHWRTLRLRDALGEPVAPERTVFERAVREGNLASDLYRLEGSGAQPRV 460

Qy         88 --------KTAEGNPTGGLVGLRIPTSK 107
                     :| ||   ||:  ||  |::
Db        461 LQITAAPLRTGEGELLGGVAVLRDITAQ 488


Search completed: February 28, 2006, 08:53:06
Job time : 41.609 secs
```